

(3)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number:

0 532 167 A2

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: **92306952.0**

(51) Int. Cl.⁵ **C12N 15/51, A61K 39/29,
G01N 33/576, C07K 15/00,
C12P 21/08**

(22) Date of filing: **30.07.92**

(30) Priority: **09.08.91 JP 287402/91**
05.12.91 JP 360441/91

(43) Date of publication of application:
17.03.93 Bulletin 93/11

(84) Designated Contracting States:
BE CH DE DK FR GB IT LI NL

(71) Applicant: **IMMUNO JAPAN INC.**
4-28-14-701, Ogikubo Suginami-Ku
Tokyo 167(JP)

(72) Inventor: **Okamoto, Hiroaki**
3132-24, Yakushiji, Minami Kawachi Machi
Kawachi-gun; Tochigi-ken 329-04(JP)
Inventor: **Nakamura, Tetsuo**
14-701, 4-chome 28, Ogikubo
Suginami-ku, Tokyo 167(JP)

(74) Representative: **Arthur, Bryan Edward**
4 Dyers Buildings Holborn
London, EC1N 2JT (GB)

(54) **Non-A, Non-B Hepatitis virus genome, polynucleotides, polypeptides, antigen, antibody and detection systems.**

(57) **Non-A, non-B hepatitis (NANB hepatitis) virus RNA and its corresponding polypeptide, related antigen, antibody, and detection systems for detecting NANB hepatitis antigen or antibodies.**

EP 0 532 167 A2

Reference To A Related Application

The present application is a continuation-in-part of our copending U.S. Patent Application Serial No. 07 866,045, filed on April 9, 1992, which is incorporated by reference in its entirety.

Background of the Invention

The present invention concerns non-A, non-B hepatitis (hereinafter called NANB hepatitis) virus genome, polynucleotides, polypeptides, related antigen, antibody and detection systems for detecting NANB antigens or antibodies.

Viral hepatitis of which DNA and RNA of the causative viruses have been elucidated, and their diagnosis and even prevention in some have been established, are hepatitis A and hepatitis B. The general name NANB hepatitis was given to the other forms of viral hepatitis.

Post-transfusion hepatitis was remarkably reduced after introduction of diagnostic systems for screening hepatitis B in transfusion bloods. However, there are still an estimated 280,000 annual cases of post-transfusion hepatitis caused by NANB hepatitis in Japan.

NANB hepatitis viruses were recently named C.D and E according to their types, and scientists started a world wide effort to conduct research for the causative viruses and subsequent extermination of the causative viruses.

In 1988, Chiron Corp. claimed that they had succeeded in cloning RNA virus genome, which they termed hepatitis C virus (hereinafter called HCV), as the causative agent of NANB hepatitis and reported on its nucleotide sequence (British Patent 2,212,511 which is the equivalent of European Patent Application 0,318,216). HCV (C100-3) antibody detection systems based on the sequence are now being introduced for screening of transfusion bloods and for diagnosis of patients in Japan and in many other countries. The detection systems for the C100-3 antibody have proven their partial association with NANB hepatitis; however, they capture only about 70% of carriers and chronic hepatitis patients, or they fail to detect the antibody in acute phase infection, thus leaving problems yet to be solved even after development of the C100-3 antibody by Chiron Corp.

The course of NANB hepatitis is troublesome and most patients are considered to become carriers, then to develop chronic hepatitis. In addition, most patients with chronic hepatitis develop liver cirrhosis, then hepatocellular carcinoma. It is therefore very imperative to isolate the virus itself and to develop effective diagnostic reagents enabling earlier diagnosis.

The presence of a number of NANB hepatitis which cannot be diagnosed by Chiron's C100-3 antibody detection kits suggests a possibility of a difference in subtype between Chiron's HCV and Japanese NANB hepatitis virus.

In order to develop NANB hepatitis diagnostic kits of more specificity and to develop effective vaccines, it becomes an absolutely important task to analyze each subtype of NANB hepatitis causative virus at its genetic and corresponding amino acid level.

Summary of the Invention

An object of the present invention is to provide the nucleotide sequence coding for the structural protein of NANB hepatitis virus and, with such information, to analyze amino acids of the protein to locate and provide polypeptides useful as antigen for establishment of detection systems for NANB virus, its related antigens and antibodies.

A further object of the present invention is to locate polynucleotides essential to treatment, prevention and diagnosis, and polypeptides effective as antigens, by isolating NANB hepatitis virus RNA from human and chimpanzee virus carriers, cloning the cDNA covering the whole structural gene of the virus to determine its nucleotide sequence, and studying the amino acid sequence of the cDNA. As a result, the inventors have determined the nucleotides of the whole genome of a strain of NANB virus called HC-J6 and a strain called HC-J8. NANB hepatitis virus genome of HC-J6 and HC-J8 differ from that of Chiron's HCV.

Brief Description of the Drawings

Figure 1 shows the restriction map and structure of the coding region of NANB hepatitis virus genome (HC-J6) and positions of clones. C, E, NS-1, NS-2, NS-3, NS-4 and NS-5 are the abbreviation of core, envelope, non-structure-1, -2, -3, -4 and -5.

Figures 2 to 4 show method of determination of the nucleotide sequence of 5' terminus of NANB hepatitis virus genome of strains HC-J1, HC-J4 and HC-J6 respectively.

Figure 5 shows the method of determination of the nucleotide sequence of 3' terminus of HC-J6 genome. Solid lines show nucleotide sequences determined by clones from libraries of bacteriophage lambda gt10, and broken lines show nucleotide sequences determined by clones obtained by PCR.

Figure 6 shows the structure of coding region of NANB hepatitis virus genome (HC-J8) and positions of clones. Regions a to n indicate positions of amplification by PCR.

Detailed Description of the Invention

10

The present invention provides NANB hepatitis virus genome RNA for strain HC-J6 (sequence list 1) consisting of 340 nucleotides on the 5' terminus that follow an open reading frame consisting of 9099 nucleotides coding for the structural protein and non-structural protein that follow a noncoding region consisting of 150 nucleotides containing an U-stretch consisting of 108 uracils on the 3' terminus of NANB hepatitis virus, and NANB hepatitis virus genome having substantially the nucleotide sequence of sequence list 1.

The present invention provides polynucleotide N-9589 (strain HC-J6) comprising the DNA nucleotide sequence of sequence list 2; cDNA clone J6-ø81 comprising the nucleotide sequence of sequence list 3; cDNA clone J6-ø8 comprising the nucleotide sequence of sequence list 4; and NANB hepatitis virus polynucleotides having substantially the sequence of nucleotides of NANB hepatitis virus nucleotides shown in sequence lists 2 through 4.

The invention provides polypeptide coded for by genome or polynucleotide of HC-J6 above, polypeptide P-J6-3033, comprising the polypeptide sequence of sequence list 5, polypeptides produced by using recombinant genome, recombinant polynucleotides and recombinant cDNA of whole or a part of cDNA above, and polyclonal or monoclonal antibodies against the polypeptides described above.

The present invention also provides NANB hepatitis virus genome for strain HC-J8 comprising sequence list 6, NANB hepatitis virus RNA consisting of noncoding region consisting of 341 nucleotides on 5' terminus followed by an open reading frame consisting of 9099 nucleotides coding for the structural protein and non-structural protein followed by a noncoding region consisting of 71 nucleotides containing an U-stretch consisting of 30 uracils on 3' terminus of NANB hepatitis virus comprising sequence list 6, and NANB hepatitis virus genome having substantially the nucleotide sequence of sequence list 6.

The present invention provides polynucleotide N-9511 for strain HC-J8 comprising the DNA nucleotide sequence of sequence list 7 and NANB hepatitis virus polynucleotide having substantially the sequence of nucleotides of NANB hepatitis virus nucleotides comprising sequence list 7.

The invention provides polypeptide coded for by genome or polynucleotide of HC-J8 above, polypeptide P-J8-3033, comprising the polypeptide sequence of sequence list 8 and polypeptide P-J8-3033-2 comprising the polypeptide sequence of sequence list 9, polypeptides produced by using recombinant genome, recombinant polynucleotides and recombinant cDNA of whole or a part of cDNA above, and polyclonal or monoclonal antibodies against the polypeptides described above.

The present invention, furthermore, provides NANB hepatitis diagnostic system using polypeptides or antibodies described above.

In the method described below, NANB hepatitis virus RNA of the present invention was obtained and its nucleotide sequence was determined.

Plasma samples (HC-J1, HC-J4, HC-J6 and HC-J8) were obtained from human and chimpanzee. HC-J1, HC-J6 and HC-J8 were obtained from Japanese blood donors who had tested positive for HCV antibody. HC-J4 was obtained from the chimpanzee subjected to the challenge test but was negative for Chiron's C100-3 antibody previously mentioned.

RNA was isolated from each of the plasma samples. Following the study of 5' terminus of approximately 2,500 nucleotides and 3' terminus of approximately 1,100 nucleotides disclosed in Japanese patent application No. 196175/91, the inventors have completed the study of the region coding for non-structural protein of strain HC-J6 and the study of the full length sequence of 9,589 nucleotides of HC-J6 genome RNA and have completed the study of the region coding for non-structural protein of strain HC-J8 and the study of the full length sequence of 9,589 nucleotides of HC-J8 genome RNA.

As described in the Example below, strain HC-J6 had a 5' noncoding region consisting of 340 nucleotides, and strain HC-J8 had a 5' noncoding region consisting of 341 nucleotides, followed by region coding for structural protein and region coding for non-structural protein.

Concerning the 3' terminus, strain HC-J6 was found to have a region consisting of 150 nucleotides containing an U-stretch consisting of 108 uracils following after the region coding for non-structural protein

and strain HC-J8 was found to have a region consisting of 71 nucleotides containing an U-stretch consisting of 30 uracils following after the region coding for non-structural protein.

The coding region starting with adenine (341st nucleotide from the 5' terminus for strain HC-J6 and 342nd nucleotide from the 5' terminus for strain HC-J8) was found to have a long *Open Reading Frame* consisting of 9099 nucleotides which codes for 3033 amino acids. HCV or hepatitis C virus is supposed to be closely allied to flavivirus in regard to its genetic structure. The coding of the NANB hepatitis virus genome of the present invention was considered to be consisting of regions named C (core), E (envelope), NS-1 (non-structural-1), NS-2 (non-structural-2), NS-3 (non-structural-3), NS-4 (non-structural-4) and NS-5 (non-structural-5).

As compared with the sequence of HCV disclosed in the European Patent Application by Chiron Corp. (Publication No. 388,232), homology of sequences of the strain HC-J6 was 67.9% for the full nucleotide sequence and 72.3% for the full amino acid sequence, and homology of sequences of the strain HC-J8 was 66.4% for the full nucleotide sequence and 71.0% for the full amino acid sequence.

From an examination of homology for regions, the homology of nucleotide sequences (strain HC-J6) of the 5' terminal noncoding region was 94.4% and that of the amino acid sequences of the C region was 90.1%, showing comparatively high homology; on the other hand, concerning lower stream than envelope, homologies of amino acid sequence were found to be as low as 60.4% for E, 71.1% for NS-1, 57.8% for NS-2, 81.1% for NS-3, 73.1% for NS-4, and 69.9% for NS-5. As a result, HC-J6 strain was found to be significantly different from HCV strain found by Chiron Corp.

From an examination of homology for regions, the homology of nucleotide sequences (strain HC-J8) of the 5' terminal noncoding region was 93.8% and that of the amino acid sequences of the C region was 90.1%, showing comparatively high homology; on the other hand, concerning lower stream than envelope, homologies of amino acid sequence were found to be as low as 54.7% for E, 73.1% for NS-1, 55.6% for NS-2, 81.3% for NS-3, 72.1% for NS-4, 67.3% for NS-5, and 25.9% for 3' terminal noncoding region. As a result, HC-J8 strain was found to be significantly different from HCV strain found by Chiron Corp.

From the comparison of amino acid sequence of HC-J6 strain with strain HC-J1 (American type) and strain HC-J4 (Japanese type) disclosed by the inventors (Japan. J. Exp. Med. (1990), 60: 167-177), homology in the core region was more than 90% for each strain while that in the envelope region was 60.9% for HC-J1 and 53.1% for HC-J4. Thus, in the present invention, strain HC-J6 was found to be a different type of virus than strains HC-J1 or HC-J4.

From the comparison of amino acid sequence of HC-J8 strain with strain HC-J1 (type I) and strain HC-J4 (type II), homology of approximately 3,000 nucleotides of 5' terminus was 70.1% for HC-J1 and 67.1% for HC-J4, and from the comparison of all nucleotides with HC-J6 (type III) genome homology was as low as 76.9%. On the other hand, HC-J8 showed high homology with strain HC-J7 (type IV) disclosed in Japanese patent application 196175/91 as 93.1% for approximately 3,000 nucleotides of 5' terminus.

Nucleotides among strains assumed to belong to same type were supposed to show high homology. For example, homology of 95.6% for approximately 3,000 nucleotides of 5' terminus between HCV disclosed by Chiron Corp. and HC-J1 appears to show that they should be classified into type I. On the other hand, low homology of HC-J8 with HCV, HC-J1, HC-J4 and HC-J6 appeared to show that it was not to be classified into type I, II or III, but into type IV (the same as HC-J7).

Strain HC-J8 has some mutations in the nucleotides as shown in sequence lists 6 and 7 by symbols M, R, W, S, Y, K and B. It also can be easily understood that it has some mutations of amino acids from comparison of sequences in sequences lists 8 and 9. Mutation of nucleotides was observed up to approximately 1.4% in the whole genome and that of amino acids was observed up to approximately 1.7% in whole ORF. Thus the present invention includes genomes, polynucleotides and polypeptides of strain HC-J8 having some mutations.

In addition, envelope (E) region (576 nucleotides/192 amino acids of amino acids 192-383) and NS-1 region (1050 nucleotides/350 amino acids of amino acids 384-733) having many mutations in HC-J8 are called hyper-variable region since mutations were observed as 20 nucleotides/7 amino acids (3.47%/3.64%) in E region and 37 nucleotides/19 amino acids (3.52%/5.42%) in NS-1 region. According to these findings, the present invention can be recognized to include genomes and polypeptides coded for by the genomes of strain HC-J8 having mutations of 3.5% to 5.5% in those regions.

The genome, polynucleotide, and cDNA clones of the present invention can be used as material to produce peptides of the invention by integration into a host genome, e.g. *E. coli* or *Bacillus*, by means of known genetic engineering techniques.

Polypeptides of the invention are useful as material for diagnostic agents to detect NANB hepatitis antibodies with high specificity and as material to produce polyclonal and monoclonal antibodies by known techniques.

Polyclonal and monoclonal antibodies of the invention are useful as materials for diagnostic agents to detect NANB hepatitis antigens with high specificity.

A detection system using each polypeptide of the present invention or polypeptide with partial replacement of amino acids, and a detection system using monoclonal or polyclonal antibodies to such polypeptides, are useful as diagnostic agents of NANB hepatitis with high specificity and are effective to screen out NANB hepatitis virus from transfusion bloods or blood derivatives. The polypeptides, or antibodies to such polypeptides, can be used as a material for a vaccine against NANB hepatitis virus.

It is well known in the art that one or more nucleotides in a DNA sequence can be replaced by other nucleotides in order to produce the same protein. The present invention also concerns such nucleotide substitutions which yield DNA sequences which code for polypeptides as described above. It is also well known in the art that one or more amino acids in an amino acid sequence can be replaced by equivalent other amino acids, as demonstrated by U.S. Patent No. 4,737,487 which is incorporated by reference, in order to produce an analog of the amino acid sequence. Any analogs of the polypeptides of the present invention involving amino acid deletions, amino acid replacements, such as replacements by other amino acids, or by isosteres (modified amino acids that bear close structural and spatial similarity to protein amino acids), amino acid additions, or isosteres additions can be utilized, so long as the sequences elicit antibodies recognizing NANB antigens.

Examples of application of this invention are shown below, however, the invention shall in no way be limited to those examples

Examples

The 5' terminal nucleotide sequence and amino acid sequence of NANB hepatitis virus genome were determined in the following way:

(1) Isolation of RNA

RNA of the sample (HC-J1, HC-J6, HC-J8) from plasma of Japanese blood donor testing positive for HCV (C100-3) antibody (by Ortho HCV Ab ELISA, Ortho Diagnostic System, Tokyo), and that of the sample (HC-J4) from the chimpanzee challenged with NANB hepatitis for infectivity and negative for HCV antibody were isolated in the following method:

Each plasma sample was added with Tris chloride buffer (10 mM, pH 8.0) and centrifuged at 68×10^3 rpm for 1 hour. Its precipitate was suspended in Tris chloride buffer (50 mM, pH 8.0) containing 200 mM NaCl, 10 mM EDTA, 2% (w/v) sodium dodecyl sulfate (SDS), and proteinase K 1 mg/ml, incubated at 60°C for 1 hour, then their nucleic acids were extracted by phenol/chloroform and precipitated by ethanol to obtain RNA.

(2) HC-J1 and HC-J8 cDNA Synthesis

After heating the RNA isolated from HC-J1 or HC-J8 plasma at 70°C for 1 minute, this was used as a template; 10 units of reverse transcriptase (cDNA Synthesis System Plus, Amersham Japan) and 20 pmol of oligonucleotide primer (20 mer) were added and incubated at 42°C for 1.5 hours to obtain cDNA. Primer #8 (5'- GATGCTTGCGGAAGCAATCA - 3') was prepared by referring to the basic sequence shown in European Patent Application No. 88310922.5, which is relied on and incorporated herein by reference.

(3) cDNA Was Amplified by the following Polymerase Chain Reaction (PCR)

cDNA was amplified for 35 cycles according to Saiki's method (Science (1988) 239: 487-491) using Gene Amp DNA Amplifier Reagent (Perkin-Elmer.Cetus) on a DNA Thermal Cycler (Perkin-Elmer.Cetus).

For cDNA synthesis and for PCR for HC-J8, synthesized primers disclosed in Japanese patent application 153402/90 and those based on HC-J1, HC-J4 and HC-J6 genomes disclosed in Japanese patent applications 196175/91 and below were utilized.

(4) Determination of 5' Terminal Nucleotide Sequence of HC-J1 and HC-J4 by Assembling cDNA Clones

As shown in Figures 2 and 3, nucleotide sequences of 5' termini of the genomes of strains HC-J1 and HC-J4 were determined by combined analysis of clones obtained from the cDNA library constructed in bacteriophage λ gt10 and clones obtained by amplification of HCV specific cDNA by PCR.

Figures 2 and 3 show 5' termini of NANB hepatitis virus genome together with cleavage site by restriction endonuclease and sequence of primers used. In the figures, solid lines are nucleotide sequences determined by clones from bacteriophage λ gt10 library while dotted lines show sequences determined by clones obtained by PCR.

A 1656 nucleotide sequence of HC-J1 spanning nt454-2109 was determined by clone ϕ 41 which was obtained by inserting the cDNA synthesized with the primer #8 into λ gt10 phage vector (Amersham).

Another primer #25 (5'- TCCCTGTTGCATAGTTCACG -3') corresponding to nt824-843 was synthesized based on the ϕ 41 sequence, and four clones (ϕ 60, ϕ 61, ϕ 66 and ϕ 75) were obtained to cover the upstream sequence nt18-843.

(5) Determination of 5' Terminal Nucleotide Sequence of HC-J6.

The nucleotide sequence of the 5' terminus of strain HC-J6 was determined from analysis of clones obtained by PCR amplification as shown in Figure 4.

Isolation of RNA from HC-J6 and determination of its sequence was made in the same manner as described in (2) above. Sequences in the range of nt24-2551 of the RNA were determined from consensus sequence of respective clones obtained by amplification by PCR using each pair of primers based on nucleotide sequence of HC-J4.

nt24-826

#32 (5'-ACTCCACCATAGATCACTCC-3')

#122 (5'-AGGT'TCCCTGTTGCATAATT-3')

Clones: C9397, C9388, C9764

nt732-1907

#50 (5'-GCCGACCTCATGGGGTACAT-3')

#128 (5'-TCGGTCGTGCCCCACTACCAC-3')

Clones: C9316, C9752, C9753

nt1847-2571

#149 (5'-TCTGTGTGTGGCCCAGTGTA-3')

#146 (5'-AGTAGCATCATCCACAAGCA-3')

Clones: C11621, C11624, C11655

In order to determine further upstream of the 5' terminus, antisense primer #36 (5'- AACACTACTCGG-CTAGCAGT -3') corresponding to nt246-265, followed by dAs were added to 5' terminus of cDNA using terminal deoxynucleotidyl transferase, and one-sided PCR amplification was made twice as described below.

cDNA was amplified for 35 cycles as first stage PCR using oligo dT primer (20-mer) and antisense primer #48 (5'-GTTGATCCAAGAAAGGACCC -3') of nt188-207, followed by the second stage of PCR by 30 cycle amplification using the first PCR product as a template, oligo dT primer (20 -mer) and antisense

primer #109 (21-mer; 5'-ACCGGATCCGCAGACCACTAT-3') corresponding to nt140 to 160. The obtained PCR product was subcloned to M13 phage vector.

Nucleotide sequence from nt1 to 23 was determined from consensus sequence of 13 isolated clones C9577, C9579, C9581, C9587, C9590, C9591, C9595, C9606, C9609, C9615, C9616 and C9619 obtained above which were considered having complete 5' terminus.

(6) Determination of nucleotide sequence of HC-J6 middle region.

cDNA library was constructed with using λ gt10 according to the method described in (2) above from 100ml of HC-J6 plasma as a starting materials. Primers #162 and #81 were prepared for synthesis by referring to the basic sequence shown in the European Patent Application Publication No. 318,216. Clones were selected by plaque hybridization.

Nucleotide sequence from 2552 to 8700 was determined from consensus sequence of four obtained cDNA clones ϕ 2 (nt6996 to 8700), ϕ 6 (nt6485 to 8700), ϕ 8 (nt6008 to 8700) and ϕ 81 (nt2199 to 6168) as shown in Figure 1. Clones ϕ 81 and ϕ 8 were found to have nucleotide sequences shown in sequence lists 3 and 4 respectively.

(7) Determination of 3' terminal nucleotide sequence of HC-J6 strain.

As shown in Figure 5, the nucleotide sequence of the 3' terminus of HC-J6 genome was determined by analysis of clones obtained by amplification of HCV specific cDNA by PCR.

Nucleotide sequence of HC-J6 from nt8701 to 9241 was determined from consensus sequence of three clones consisting of 938 nucleotides, C9760, C9234 and C9761, obtained by amplification of sample using primer #80 (5'-GACACCCGCTGTTTTGACTC-3') and #60 (5'-GTTCTTACTGCCAGTTGAA-3').

Nucleotide sequence of 3' terminus down stream from nt9242 was determined in the method described below.

Isolation of RNA from HC-J6 was made in the same manner as described in (1) above. The obtained RNA was added poly (A) to its 3' terminus using poly (A) polymerase and cDNA was synthesized using oligo (dT)₂₀ as a primer, and obtained cDNA was provided to PCR as a template.

First PCR product was made with using #97 (5'-AGTCAGGGCGTCCCTCATCT-3') as a sense primer and oligo (dT)₂₀ as an antisense primer. Second PCR product was made with using #90 (5'-GCCGTTTGCGCCGATATCT-3') corresponding to downstream sequence of #97 as a sense primer, and oligo (dT)₂₀ as an antisense primer as well as first PCR product. PCR product obtained by two step amplification was smoothened on both ends by treatment with T₄ DNA polymerase, followed by phosphorylation of 5' terminus by T₄ polynucleotide kinase. The obtained product was subcloned into Hinc II position of M13mp19 phage vector.

Nucleotide sequence of 3' terminus was determined from consensus sequence of 19 obtained clones, C10311, C10313, C10314, C10320, C10322, C10323, C10326, C10328, C10330, C10333, C10334, C10336, C10337, C10345, C10346, C10347, C10349, C10350 and C10357.

As a result, the nucleotide sequence of cDNA to HC-J6 genome RNA was determined as shown in sequence list 2, and full sequence of genome RNA was determined as shown in sequence list 1.

(8) Determination of amino acid sequences.

According to the nucleotide sequence of the genome of strain HC-J6, determination was made of sequence of coded region starting with ATG. As a result, HC-J6 genome was found to have a long *Open Reading Frame* coding for polypeptide precursor consisting of 3033 amino acid residues.

(9) Determination of 5' terminal nucleotide sequence of HC-J8

As shown in Figure 6, the nucleotide sequence of 5' terminus of HC-J8 genome (a region) was determined by analysis of clones obtained by amplification of HCV specific cDNA by PCR.

Single-stranded cDNA was synthesized using antisense primer #36 (5'-AACACTACTCGGCTAGCAGT-3') of nt246 to 265 in the same manner as (2) above, then it was added with dATP tail at its 3' terminus by terminal deoxynucleotidyl transferase, then amplified by one-sided PCR in two stages.

That is, in the first stage, antisense primer #48 (5'-GTTGATCCAAGAAAGGACCC-3') of nt188 to 207 was used with sense primer selected from non-specific primer #165 (5'-AAGGATCCGTCGACATCGATAAT-ACG (A)_{17-3'}) and #171 (5'-AAGGATCCGTCGACATCGATAATACG(T)_{17-3'}) to amplify the dA-tailed cDNA

by PCR for 35 cycles; and in the second stage, using the product of the first-stage PCR as a template, non-specific primer #166 (5' AAGGATCCGTCGACATCGAT -3') and antisense primer #109 (21-mer, 5'-ACCGGATCCGCAGACCACTAT -3') were added to initiate PCR for 30 cycles. The product of PCR was subcloned to M13 phage vector.

5 Thirteen independent clones (poly dT-tailed: C14951, C14952, C14953, C14958, C14960, C14968, C14971, C14972 and C14974; poly dA-tailed: C14987, C14996, C14999 and C15000) were obtained (each considered having complete length of 5' terminus), and the consensus sequence of nt1-139 of the respective clones was determined.

10 (10) cDNA amplification of ORF region and 3' terminus by PCR

As shown in Figure 6, the nucleotide sequence of downstream from nt140 of HC-J8 genome was determined by analysis of clones obtained by amplification of HCV specific cDNA by PCR.

15 Single-stranded cDNAs to HC-J8 RNA were synthesized in the same manner as (2) above using antisense primers described below, then they were amplified by PCR using sense and antisense primers described below. Each product of PCR was subcloned to M13 phage vector, then consensus sequence of the respective clones of each region was determined.

The primers for cDNA synthesis and PCR amplification, and the numbers of obtained clones are shown below for each region. Alphabetical symbol of each amplified region corresponds to that in Figure 6.

20

25

30

35

40

45

50

55

b region

nt45-847

5 Primer for cDNA synthesis: #122 (5'-AGGTTCCCTGTTGCATAATT-3')

Primer for PCR: sense: #32A (5'-CTGTGAGGAACTACTGTCTT-3')

10 antisense #122

Clones: C15221,C15222,C15223

15 c region

nt732-1354

20 Primmer for cDNA synthesis:#54 (5'-ATCGCGTACGCCAGGATCAT-3')

Primer for PCR: sense: #50 (5'-GCCGATCTCATGGGGTACAT-3')

antisense:#54

25 Clones: C15256,C15257,C15258

30 d region

ntl300-1879

Primer for cDNA synthesis: #199 (5'-GGGGTGAAACAATACACCGG-3')

35 Primer for PCR: sense: #205 (5'-GGGACATGATGATCAACTGG-3')

antisense: #199

40 Clones: C14221,C14222,C14223

45

50

55

e region

ntl833-2518

Primer for cDNA synthesis: #146 (5'-AGTAGCATCATCCACAAGCA-3')

Primer for PCR: sense: #150 (5'-ATCGTCTCGGCTAAGACGGT-3')

antisense: #146

Clones: C11535, C11540, C11566

f region

nt2433-3451

Primer for cDNA synthesis: #170 (5'-GCATAAGCAGTGATGGGGGC-3')

Primer for PCR: sense: #160 (5'-CAGAACATCGTGGACGTGCA-3')

antisense: #170

Clones: C15348, C15349, C15356

g region

nt3404-4300

Primer for cDNA synthesis: #225 (5'-TCGCATATGATGATGTCATA-3')

Primer for PCR: sense: #238 (5'-CTACACCTCCAAGGGGTGGA-3')

antisense: #225

Clones: C15701, C15702, C15703

h region

nt4221-5015

Primer for cDNA synthesis: #216 (5'-GTGGTCTAGACATACGGGCA-3')

Primer for PCR: sense: #230 (5'-CCCATCACGTACTCCACATA-3')

antisense: #216

Clones: C15391, C15392, C15393

5 i region

nt4695-5062

Primer for cDNA synthesis: #210 (5'-GCATCTATGTGTGTGAGGCC-3')

10 Primer for PCR: sense: #209 (5'-TTCGACTCCGTGATCGACTG-3')
antisense: #210

15 Clones: C14087, C14088, C14089

20 j region

nt5021-6169

Primer for cDNA synthesis: #162 (5'-TCCGACTCCGTCACGTAGTG-3')

25 Primer for PCR: sense: #227 (5'-GTTCTGGGAAGCGGTCTTTA-3')
antisense: #162

30 Clones: C15421, C15422, C15423

k region

35 nt6027-6889

Primer for cDNA synthesis: #232 (5'-GATGGGTCTGTAGCATGGA-3')

Primer for PCR: sense: #242 (5'-TTGGTAGTGGGAGTCATCTG-3')
40 antisense: #232

Clones: C15733, C15734, C15735

45 l region

nt6834-7735

50 Primer for cDNA synthesis #239 (5'-ATCGGTAACCTTCCTCTTC-3')

55

Primer for PCR: sense: #241 (5'-CCTTGCGATCCTGAACCTGA-3')
 antisense: #239

Clones: C15798, C15799, C15800

m region

nt7656-8630

Primer for cDNA synthesis: #222 (5'-GACCAGGTCGTCTCCACACA-3')

Primer for PCR: sense: #229 (5'-GTCGTGTGCTGCTCCATGTC-3')
 antisense: #222

Clones: C15376, C15378, C15381

n region

nt8325-9511

Primer for cDNA synthesis: #165

Primer for PCR: sense: #80 (5'-GACACCCGCTGTTTTGACTC-3')
 non-specific: #165

Clones: C15270, C15271, C15272

From the analysis described above, full nucleotide sequence of cDNA to HC-J8 was determined as shown in sequence list 7, then full nucleotide sequence of HC-J8 genome RNA as shown in sequence list 6. Two amino acid sequences shown in sequence lists 8 and 9 represent those coded for by HC-J8 genome.

Utilizing known immunological techniques, it is possible to determine epitopes (e.g., from the core region, etc.) from the polypeptides of sequence lists 5, 8 and 9. Determination of such epitopes of the NANB hepatitis virus opens access to chemical synthesis of the peptide, manufacturing of the peptide by genetic engineering techniques, synthesis of the polynucleotides, manufacturing of the antibody, manufacturing of NANB hepatitis diagnostic reagents, and development of products such as NANB hepatitis vaccines.

According to the well-known method described by Merrifield, NAMB peptides can be synthesized. Furthermore, the polynucleotides in sequence lists 2-4 and 7 can be used to express polypeptides in host cells such as *Escherichia coli* by means of genetic engineering technique.

A detection system for antibody against NANB hepatitis virus can be developed using polyvinyl microtiter plates and the sandwich method. For example, 50 μ l of 5 μ g/ml concentration of a NANB peptide can be dispensed in each well of the microtiter plates and incubated overnight at room temperature for consolidation. The microplate wells can be washed five times with physiological saline containing 0.05% Tween 20. For overcoating, 100 μ l of NaCl buffer containing 30% (v/v) of calf serum and 0.05% Tween 20 (CS buffer) can be dispensed in each well and discarded after incubation for 30 minutes at room temperature.

For determination of NANB antibodies in samples, in the primary reaction, 50 μ l of the CS buffer containing 30% calf serum and 10 μ l of a sample can be dispensed in each microplate well and incubated on a microplate vibrator for one hour at room temperature. After completion of the reaction, microplate wells

can be washed five times in the same way as previously described.

In the secondary reaction, as labeled antibody 1 ng of horseradish peroxidase labeled anti-human IgG mouse monoclonal antibodies (Fab' fragment: 22G, Institute of Immunology Co., Ltd., Tokyo, Japan) dissolved in 50 μ l of calf serum can be dispensed in each microplate well, and incubated on a microplate vibrator for one hour at room temperature. Wells can be washed five times in the same way. After addition of hydrogen peroxide (as substrate) and 50 μ l of O-phenyldiamine solution (as color developer) in each well, and after incubation for 30 minutes at room temperature, 50 μ l of 4M sulphuric acid can be dispensed in each well to stop further color development and for reading absorbance at 492 nm.

The cut-off level of this assay system can be set by measuring a number of donor samples with normal serum ALT (alanine aminotransferase) value of 34 Karmen unit or below and which tested negative for anti-HCV.

The present invention makes possible detection of NANB hepatitis virus infection which could not be detected by conventional determination methods, and provide NANB hepatitis detection kits capable of highly specific and sensitive detection at an early phase of infection.

These features allow accurate diagnosis of patients at an early stage of the disease and also help to remove at higher rate NANB hepatitis virus carrier bloods through screening test of donor bloods.

Polypeptides and their antibodies under this invention can be utilized for manufacture of vaccines and immunological pharmaceuticals, and structural gene of NANB hepatitis virus provides indispensable tools for detection of polypeptide antigens and antibodies.

Antigen-antibody complexes can be detected by methods known in this art. Specific monoclonal and polyclonal antibodies can be obtained by immunizing such animals as mice, guinea pigs, rabbits, goats and horses with NANB peptides (e.g., bearing NANB hepatitis antigenic epitope).

The present invention is based on studies on isolated virus genome of NANB hepatitis virus named HC-J6 and HC-J8, and is completed by clarification of the full sequence of the nucleotides. The invention makes possible highly specific detection of NANB hepatitis virus and provision of polypeptide, polyclonal antibody and monoclonal antibody to prepare the test system.

Further variations and modifications of the invention will become apparent to those skilled in the art from the foregoing and are intended to be encompassed by the claims appended hereto.

Japanese Priority Applications 287402/91 filed August 9, 1991 and 360441/91 filed on December 5, 1991 are relied on and incorporated by reference. U.S. patent applications serial no. 07/540,604 (filed June 19, 1990), 07/653,090 (filed February 8, 1991), and 07/712,875 (filed June 11, 1991) are incorporated by reference in their entirety.

Sequence list

- | | |
|------------------|---|
| Sequence list 1 | whole nucleotides of HC-J6 genome RNA |
| Sequence list 2 | N-9589 whole nucleotides of cDNA to HC-J6 genome RNA |
| Sequence list 3 | J6- ϕ 81 nucleotides of clone J6- ϕ 81 |
| Sequence list 4 | J6- ϕ 8 nucleotides of clone J6- ϕ 8 |
| Sequence list 5 | P-J6-3033 whole amino acids of ORF of HC-J6 genome |
| Sequence list 6 | whole nucleotides of HC-J8 genome RNA |
| Sequence list 7: | whole nucleotides of cDNA to HC-J8 genome RNA |
| Sequence list 8: | whole amino acids of a variation of ORF of HC-J8 genome |
| Sequence list 9: | whole amino acids of a variation of ORF of HC-J8 genome |

Claims

1. Recombinant RNA of non-A, non-B hepatitis virus, strain HC-J6, comprising the nucleotide sequence of sequence list 1.
2. Recombinant cDNA of non-A, non-B hepatitis virus, strain HC-J6, comprising the nucleotide sequence of sequence list 2.
3. cDNA clone J6- ϕ 81 comprising the nucleotide sequence of sequence list 3.
4. cDNA clone J6- ϕ 8 comprising the nucleotide sequence of sequence list 4.

5. Amino acid sequence corresponding to recombinant cDNA of non-A, non-B hepatitis virus, strain HC-J6, comprising the amino acid sequence of sequence list 5.
6. Recombinant RNA of non-A, non-B hepatitis virus, strain HC-J8, comprising the nucleotide sequence of sequence list 6.
7. Recombinant cDNA of non-A, non-B hepatitis virus, strain HC-J8, comprising the nucleotide sequence of sequence list 7.
8. Amino acid sequence corresponding to recombinant cDNA of non-A, non-B hepatitis virus, strain HC-J8, comprising the amino acid sequence of sequence list 8.
9. Amino acid sequence corresponding to recombinant cDNA of non-A, non-B hepatitis virus, strain HC-J8, comprising the amino acid sequence of sequence list 9.
10. A non-A, non-B hepatitis diagnostic test kit for analyzing samples for the presence of antibodies directed against a non-A, non-B hepatitis antigen, comprising an antigen attached to a solid substrate and labeled anti-human immunoglobulin; wherein said antigen is an antigen selected from the antigens contained in sequence lists 5, 8 or 9.
11. A method of detecting antibodies directed against a non-A, non-B hepatitis antigen in a sample, said method comprising:
 - (a) reacting said sample with an antigen selected from the antigens contained in sequence lists 5, 8 or 9 to form antigen-antibody complexes; and
 - (b) detecting said antigen-antibody complexes.
12. A non-A, non-B hepatitis specific monoclonal or polyclonal antibody reactive with an antigen, said antigen is an antigen selected from the antigens contained in sequence lists 5, 8 or 9.
13. A method of detecting non-A, non-B hepatitis antigen in a sample, said method comprising:
 - (a) reacting said sample with the non-A, non-B hepatitis monoclonal or polyclonal antibody according to claim 12 to form antigen-antibody complexes; and
 - (b) detecting said antigen-antibody complexes.

Fig. 1

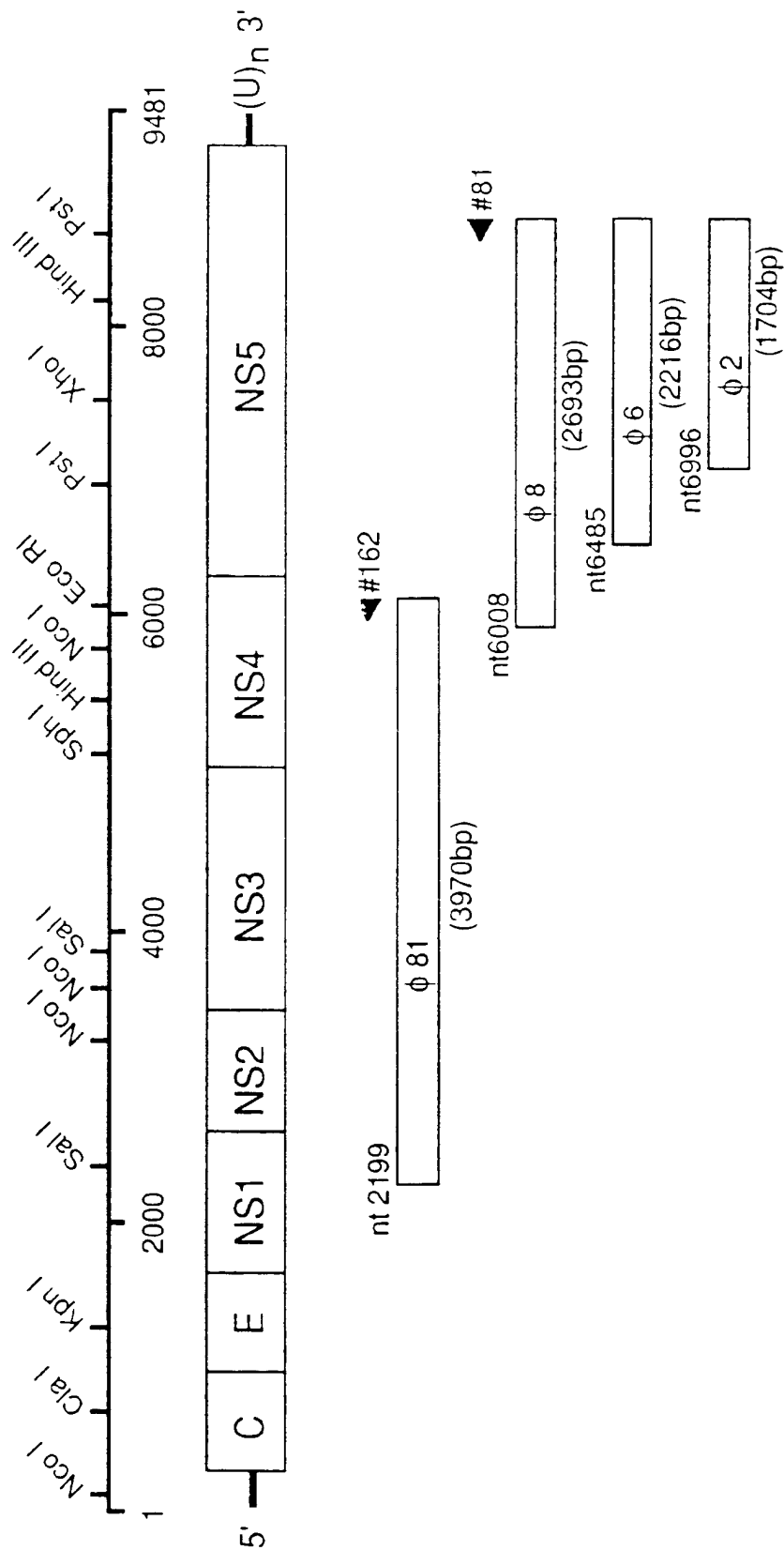


Fig. 3

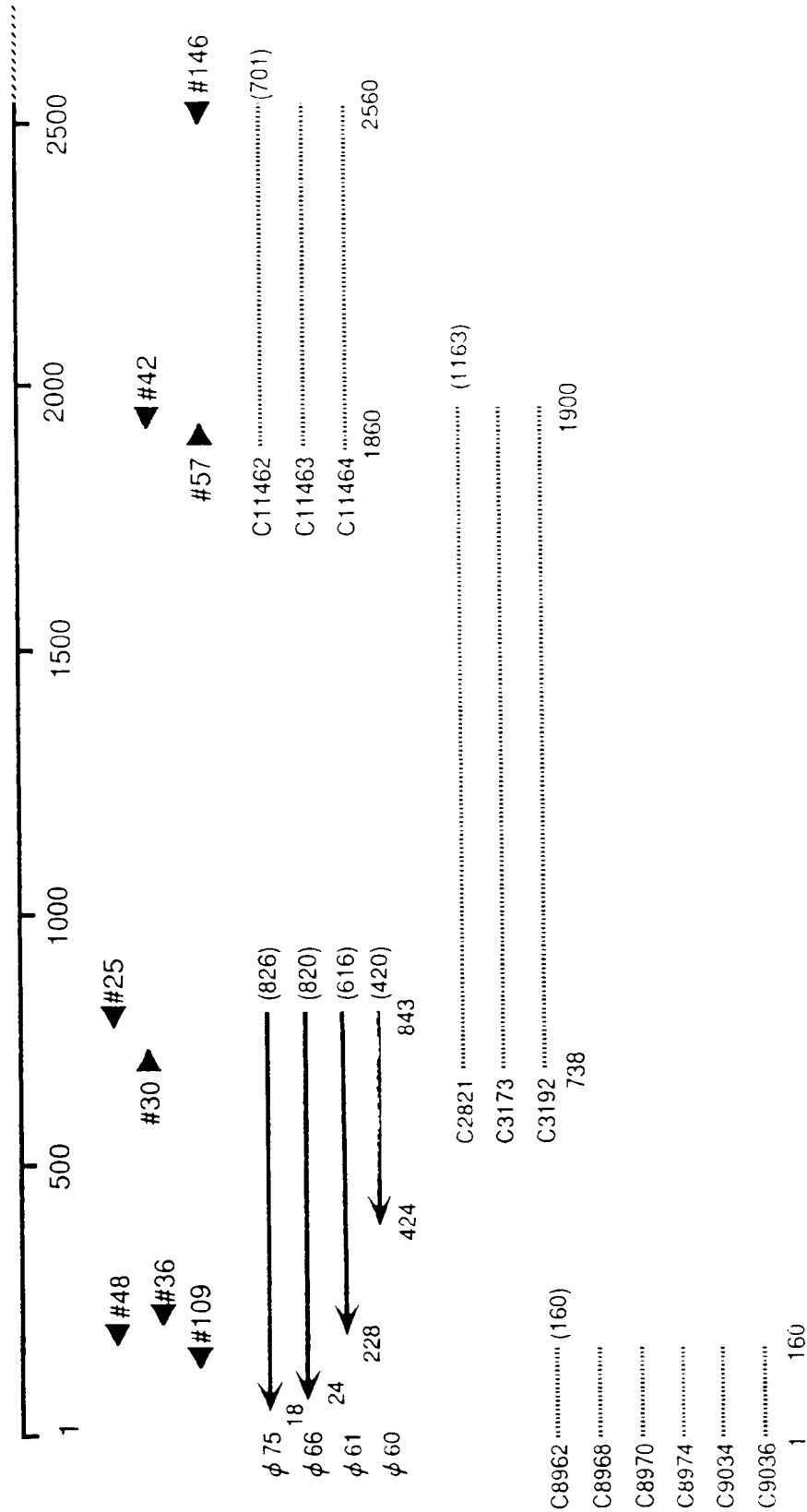


Fig. 4

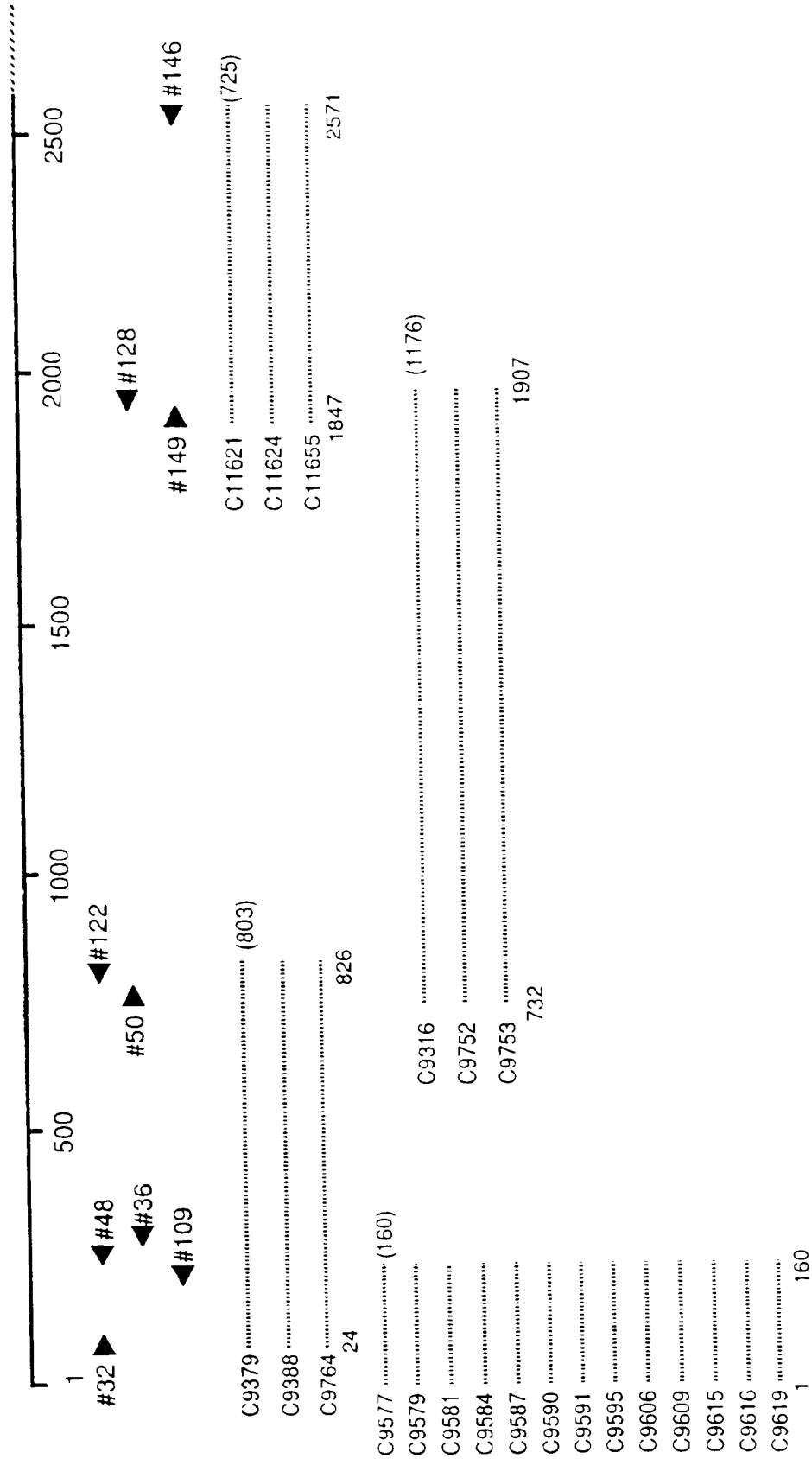


Fig. 5

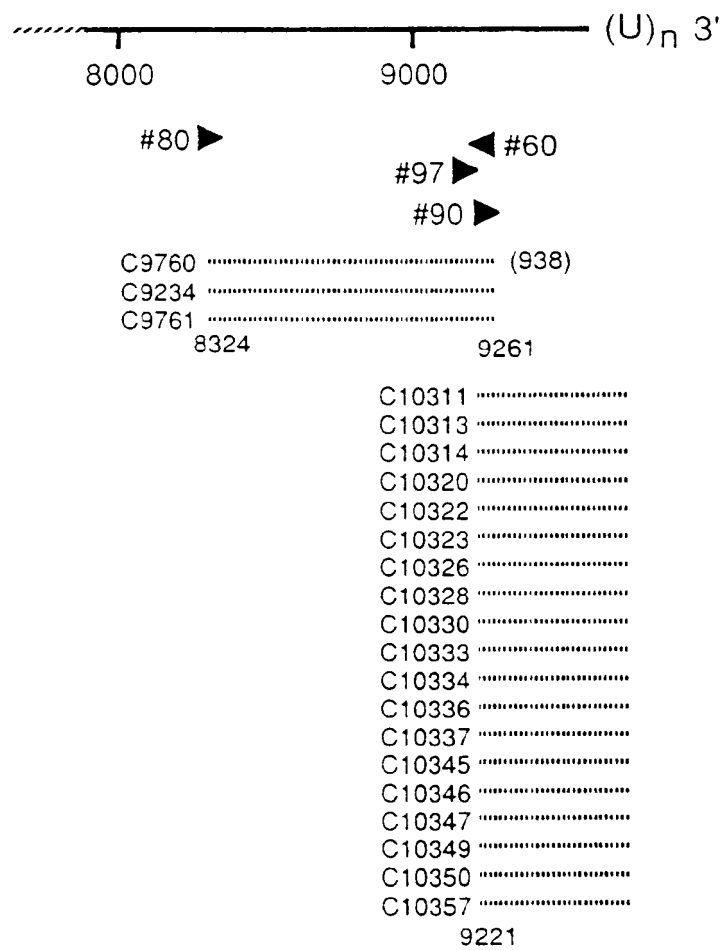
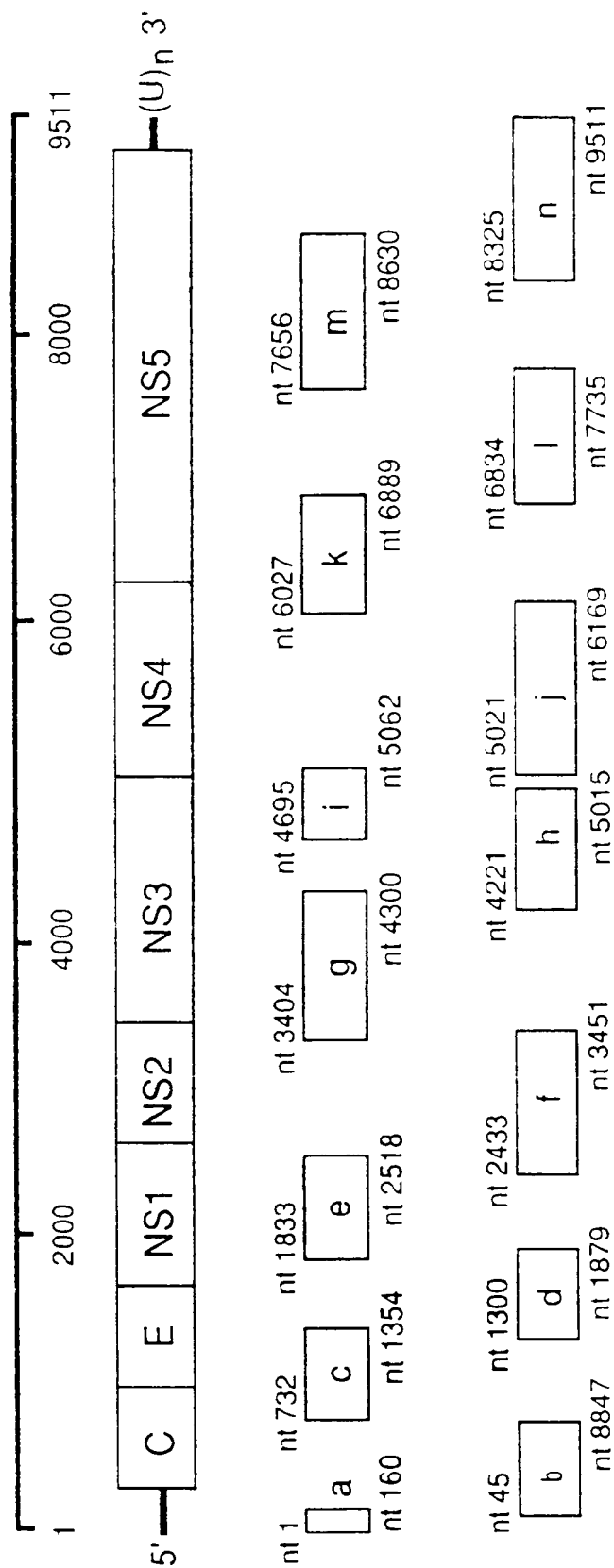


Fig. 6



Sequence ID No.1
 Sequence Length: 9,589
 Sequence Type: nucleic acid
 Strandedness: single
 Topology: linear
 Molecule Type: genomic RNA
 Method for Determination of Feature: E

```

ACCCGCCCCU AAUAGGGGCG ACACUCCGCC AUGAACCACU CCCUGUGAG GAACUACUGU   60
CUUCACGCAG AAAGCGUCUA GCCAUGGCGU UAGUAUGAGU GUCGUACAGC CUCCAGGCCC  120
CCCCCUCGCG GGAGAGCCAU AGUGGUCUGC GGAACCGGUG AGUACACCGG AAUUGCCGGG  180
AAGACUGGGU CCUUUCUUGG AUAAACCCAC UCUAUGCCCG GUCAUUUGGG CGUGCCCCCG  240
CAAGACUGCU AGCCGAGUAG CGUUGGGUUG CGAAAGGCCU UGUGGUACUG CCUGAUAGGG  300
UGCUGGCGAG UGCCCCGGGA GGUCUCGUAG ACCGUGCACC AUGAGCACA AUCCUAAACC  360
UCAAAGAAAA ACCAAAAGAA ACACCAACCG UCGCCACAAA GACGUUAAGU UUCCGGGCGG  420
CGGCCAGAUU GUUGGCGGAG UAUACUUGUU GCCGCGCAGG GGCCCCAGGU UGGGUGUGCG  480
CGCGACAAGG AAGACUUCGG AGCGGUCCCA GCCACGUGGA AGGCGCCAGC CCAUCCCUAA  540
GGAUCGGCGC UCCACUGGCA AAUCCUGGGG AAAACCAGGA UACCCUGGC CCCUAUACGG  600
GAAUGAGGGA CUCGGCUGGG CAGGAUGGCU CCUGUCCCCC CGAGGUUCCC GUCCCUCUUG  660
GGGCCCCAAU GACCCCCGGC AUAGGUCCCG CAACGUGGGU AAGGUCAUCG AUACCCUAC  720
GUGCGGCUUU GCCGACCUCA UGGGGUACAU CCCUGUCGUA GGCGCCCGC UCGGCGGCGU  780
CGCCAGAGCU CUCGCGCAUG GCGUGAGAGU CCUGGAGGAC GGGGUUAAUU UUGCAACAGG  840
GAACUUACCC GGUUGCUCCU UUUCUAUCUU CUUGCUGGCC CUGCUGUCCU GCAUCACCAC  900
CCCGGUCUCC GCUGCCGAAG UGAAGAACAU CAGUACCGGC UACAUUGGUA CCAACGACUG  960
CACCAUGAU AGCAUUACCU GGCAACUCCA GGCUGCUGUC CUCCACGUCC CCGGGUGCGU 1020
CCCGUGCGAG AAAGUGGGGA AUACAUCUCG GUGCUGGAUA CCGGUCUCAC CGAAUGUGGC 1080
CGUGCAGCAG CCCGGCGCCC UCACGCAGGG CUUACGGACG CACAUUGACA UGGUUGUGAU 1140

```

GUCCGCCACG CUCUGCUCCG CUCUUUACGU GGGGGACCUC UGCGGUGGGG UGAUGCUUGC 1200
 AGCCCAGAUG UUCAUUGUCU CGCCACAGCA CCACUGGUUU GUGCAAGACU GCAAUUGCUC 1260
 CAUCUACCCU GGUACCAUCA CUGGACACCG CAUGGCGUGG GACAUGAUGA UGAACUGGUC 1320
 GCCCACGGCU ACCAUGAUCC UGGCGUACGC GAUGCGCGUC CCCGAGGUCA UCAUAGACAU 1380
 CAUUGGCGGG GCUCAUUGG GCGUCAUGUU CGGCUUAGCC UACUUCUCUA UGCAGGGAGC 1440
 GUGGGCAAAA GUCGUUGUCA UUCUUUUGCU GGCCGCCGGG GUGGACGCGC AAACCCAUAC 1500
 CGUUGGGGGU UCUACCGCGC AUAACGCCAG GACCUCACC GGCAUGUUCU CCCUUGGUGC 1560
 CAGGCAGAAA AUCCAGCUCA UCAACACCAA UGGCAGUUGG CACAUCAACC GCACGCCCCU 1620
 GAACUGCAAU GACUCUUGC ACACGGGCUU CCUCGCGUCA CUGUUCUACA CCCACAGCUU 1680
 CAACUCGUCA GGAUGUCCCG AACGCAUGUC CGCCUGCCGC AGUAUCGAGG CCUUUCGGGU 1740
 GGAUGGGGGC GCCUUA CAU AUGAGGACAA UGUCACCAU CCAGAGGAUA UGAGACCGUA 1800
 UUGCUGGCAC UACCCACCAA GACAGUGUGG UGUAGUCUCC GCGAGCUCUG UGUGUGGCC 1860
 AGUGUACUGU UUCACCCCCA GCCCAGUAGU AGUGGGUACG ACCGAUAGAC UUGGAGCGCC 1920
 CACUACACG UGGGGGGAGA AUGAGACAGA UGUCUCCUA UUGAACAGCA CUCGACCACC 1980
 GCAGGGGUCA UGGUUCGGCU GCACGUGGAU GAACUCCACU GGUACACCA AGACUUGCGG 2040
 CGCACCACC UGCCGCAUUA GAGCUGACUU CAAUGCCAGC AUGGACUUGU UGUGCCCCAC 2100
 GGACUGUUUU AGGAAGCAUC CUGAUACCAC CUACAUCAAA UGUGGCUCUG GGCCUGGCU 2160
 CACGCCAAGG UGCCUGAUCG ACUACCCCUA CAGGCUCUGG CAUUACCCCU GCACAGUUA 2220
 CUAUACCAUC UUCAAAUAA GGAUGUAUGU GGGGGGGGUC GAGCACAGGC UCACGGCUGC 2280
 GUGCAAUUC ACUCGUGGGG AUCGUUGCAA CUUGGAGGAC AGAGACAGAA GUCAACUGUC 2340
 JCCUUGCUG CACUCCACCA CGGAGUGGGC CAUUUUACCU UGCACUACU CGGACCUGCC 2400
 CGCCUUGUCG ACUGGUCUUC UCCACCUCCA CAAAACAUC GUGGACGUGC AAUUCUUGUA 2460
 UGGCCUAUCA CCUGCUCUCA CAAAUAUUAU CGUCCGAUGG GAGUGGUAG UACUCUUAUU 2520
 CCUGCUCUUA GCGGACGCCA GGGUUUGCGC CUGCUAUGG AUGCUCAUCU UGUUGGGCCA 2580
 GGCCGAAGCA GCACUAGAGA AGUUGGUCGU CUUGCACGCU GCGAGCGCAG CUAGCUGCAA 2640
 UGGCUUCCUA UACUUGUCA UCUUUUUCGU GGCUGCUUGG UACAUCAAGG GUCGGGUAGU 2700
 CCCCUGGCU ACUUAUUCU UCACUGGCCU AUGGUCCUUU GGCCUACUGC UCCUAGCAUU 2760
 GCCCCAACAG GCUUAUGCUU AUGACGCAUC UGUACAUGGU CAGAUAGGAG CAGCUCUGUU 2820
 GGUACUGAUC ACUCUCUUA CACUCACCCC CGGGUAUAAG ACCCUUCUCA GCCGGUUUCU 2880

GUGGUGGUUG UGCUAUCUUC UGACCCUGGC GGAAGCUAUG GUCCAGGAGU GGGCACCACC 2940
 UAUGCAGGUG CGCGGUGGCC GUGAUGGGAU CAUAUGGGCC GUCGCCAUU UCUGCCCGGG 3000
 UGUGGUGUUU GACAUAAACCA AGUGGCUCUU GCGGGUGCUU GGGCCUGCUU AUCUCCUAAA 3060
 AGGUGCUUUG ACGCGUGUGC CGUACUUCGU CAGGGCUCAC GCUCUACUAA GGAUGUGCAC 3120
 CAUGGUAAGG CAUCUCGCGG GGGGUAGGUA CGUCCAGAUG GUGCUACUAG CCCUUGGCAG 3180
 GUGGACUGGC ACUUACAUCU AUGACCACCU CACCCCUAUG UCGGAUUGGG CUGCUAUUGG 3240
 CCUGCGGGAC UUGGCGGUCG CCGUGGAGCC UAUCAUCUUC AGUCCGAUGG AGAAAAAGU 3300
 CAUCGUCUGG GGAGCGGAGA CAGCUGCUUG CGGGGAUAUC UUACACGGAC UUCCCGUGUC 3360
 CGCCCGACUU GGCCGGGAGG UCCUCCUUGG CCCAGCUGAU GGCUAUACCU CCAAGGGGUG 3420
 GAGUCUUCUC GCCCCAUCA CUGCUUAUGC CCAGCAGACA CGCGGCCUUU UGGGCACCAU 3480
 AGUGGUGAGC AUGACGGGGC GCGACAAGAC AGAACAGGCC GGGGAGAUUC AGGUCCUGUC 3540
 CACGGUCACU CAGUCCUUC UCGGAACAAC CAUCUCGGGG GUCUUAUGGA CUGUCUACCA 3600
 UGGAGCUGGC AACAAGACUC UAGCCGGCUC ACGGGGUCCG GUCACACAGA UGUACUCCAG 3660
 UGCUGAGGGG GACUUAUGG GGUGGCCAG CCCCCCGGG ACCAAUUCU UGGAGCCGUG 3720
 CACGUGUGGA GCGGUCGACC UAUACCUGGU CACGCGAAAC GCUGAUGUCA UCCCGGCUCG 3780
 AAGACGCGGG GACAAGCGAG GAGCGCUACU CUCCCCGAGA CCUCUUUCCA CCUUGAAGGG 3840
 GUCCUCGGGG GGCCCGGUGC UCUGCCCCAG AGGCCACGCU GUCGGGGUCU UCCGGGCAGC 3900
 CGUGUGCUCC CGGGGCGUGG CCAAGUCCAU AGAUUUUAUC CCCGUUGAGA CACUUGACAU 3960
 CGUCACUCGG UCCCCACCU UUAGUGACAA CAGCACACCA CCUGCUGUGC CCCAAACUUA 4020
 UCAGGUCGGG UACUUAUUG CCCCAGCUGG UAGUGGAAAG AGCACCAAAG UCCCUGUCGC 4080
 GUAUGCCGCU CAGGGGUACA AAGUGCUAGU GCUUAUCCC UCGGUGGCUG CCACCCUGGG 4140
 GUUUGGGGCG UACUUGUCCA AGGCACAUGG CAUCAAUCCC AACAUUAGGA CUGGGGUCAG 4200
 GACUGUGACG ACCGGGGCGC CCAUCACGUA CUCCACAUU GGCAAUUCC UCGCCGAUGG 4260
 GGGCUGCGCA GCGGGCGCCU AUGACAUCU CAUAUGCGAU GAAUGCCAUG CCGUGGACUC 4320
 UACCACCAU CUUGGCAUCG GAACAGUCCU CGAUCAAGCA GAGACAGCCG GGGUCAGGCU 4380
 AACUGUACUG GCUACGGCUA CGCCCCCGG GUCAGUGACA ACCCCCCACC CCAACAUAGA 4440
 GGAGGUGGCC CUCGGGCAGG AGGGUGAGAU CCCCUUCUUA GGGAGGGCGA UUCCCCUGUC 4500
 AUACAUCAAG GGAGGAAGAC ACUUGAUCU CUGCCACUCA AAGAAAAAGU GUGACGAGCU 4560
 CGCGGCGGCC CUUCGGGGUA UGGGCUUGAA CGCAGUGGCA UACUACAGAG GGCUGGACGU 4620

CUCCGUAUA CCAACUCAGG GAGACGUAGU GGUCGUCGCC ACCGACGCC UCAUGACGGG 4680
 GUUUACUGGA GACUUUGACU CCGUGAUCGA CUGCAACGUA GCGGJCACUC AAGUUGUAGA 4740
 CUUCAGCUUG GACCCACAU UCACCAUAAC CACACAGACU GUCCCUCAAG ACGCUGUCUC 4800
 ACQUAGCCAG CGCCGGGGCC GCACGGGCAG GGAAGACUG GUAUUUAUA GGUAUGUUUC 4860
 CACUGGUGAG CGAGCCUCAG GAAUGUUUGA CAGUGUAGUG CUCUGCGAGU GCUACGAUGC 4920
 AGGGGCCGCA UGGUAUGAGC UCACACCAGC GGAGACCACC GUCAGGCUCA GAGCAUAUUU 4980
 CAACACACCU GGUUUGCCUG UGUGCCAAGA CCAUCUUGAG UUUUGGGAGG CAGUUUUCAC 5040
 CGGCCUCACA CACAUAGAUG CCCACUCCU UUCCCAAACA AAGCAAUCGG GGGAAAAUUU 5100
 CGCAUACUUA ACAGCCUACC AGGCUACAGU GUGCGCUAGG GCCAAAGCCC CCCCCCGUC 5160
 CUGGGACGUC AUGUGGAAGU GUUUGACUCG ACUCAAGCCC ACACUCGUGG GCCCCACACC 5220
 UCUCUGUAC CGCUUGGGCU CUGUUACCAA CGAGGUCACC CUCACGCAUC CUGUGACGAA 5280
 AUACAUCGCC ACCUGCAUGC AAGCCGACCU UGAGGUCAUG ACCAGCACGU GGGUCUAGC 5340
 UGGGGGGGUC UUGGCGGCCG UCGCCGCGUA CUGCCUGGCG ACCGGGUGUG UUUGCAUCAU 5400
 CGGCCGCUUG CACGUUAACC AGCGAGCCGU CGUUGCACCG GACAAGGAGG UCCUCUAUGA 5460
 GGUUUUGAU GAGAUGGAGG AAUGUGCCUC UAGAGCGGCU CUCAUUGAAG AGGGGCAGCG 5520
 GAUAGCCGAG AUGCUGAAGU CCAAGAUCCA AGGCUUAUUG CAGCAAGCUU CCAAACAAGC 5580
 UCAAGACUA CAACCCGCGUG UGCAGGCUUC UUGGCCAAG GUAGAGCAAU UCUGGGCCAA 5640
 ACACAUUGG AACUUCAUA GCGGCAUUA AUACCUCGCA GGACUAUCAA CACUGCCAGG 5700
 GAACCCUGCU GUAGCUUCCA UGAUGGCAUU CAGUGCCGCC CUCACCAGUC CGUUGUCAAC 5760
 UAGCACCACU AUCCUUCUA ACAUUUUGG GGGCUGGCUA GCAUCCCAA UUGCGCCUCC 5820
 CGCGGGGGCU ACCGGCUUCG UCGUCAGUGG CCUGGUGGGG GCUGCCGUAG GCAGCAUAGG 5880
 CUUGGGUAAG GUGCUGGUGG ACAUCCUGGC AGGGUAUGGU GCGGGCAUUU CGGGGGCUCU 5940
 CGUCGCAUUC AAGAUAUGU CUGGCGAGAA GCCCUCAUG GAGGAUGUUG UCAACCUGCU 6000
 GCCUGGAAUU CUGUCUCCGG GUGCCUGGU GGUGGGAGUC AUCUGCGCGG CCAUCCUGCG 6060
 CCGACACGUG GGACCGGGG AAGGCGCUGU CCAUUGGAUG AAUAGGCUA UUGCCUUUGC 6120
 UUCAGAGGA AACCACGUCG CCCCCACCA CUACGUGACG GAGUCGGAUG CGUCGCAGCG 6180
 UGUGACCCAA CUACUUGGCU CCCUACCAU AACCAGCCUG CUCAGGAGAC UCCACAACUG 6240
 GAUUACUGAA GACUGCCCCA UCCAUGCAG CGGCUCGUGG CUCCGCGAUG UGUGGGAUUG 6300
 GGUUUGCACC AUCCUAACAG ACUUUAAAA CUGGCUGACC UCCAAAUUGU UCCCAAAGAU 6360

GCCUGGUCUC CCCUUUAUCU CUUGUCAAAA GGGGUACAAG GGCUGUGGG CUGGCACUGG 6420
 UAUCAUGACC ACACGGUGUC CUUGCGGGCG CAAUAUCUCU GGCAAUGUCC GCCUGGGCUC 6480
 CAUGAGAAUU ACGGGGCCCA AAACCUGCAU GAAUAUCUGG CAGGGGACCU UCCCCAUCAA 6540
 UUGUUACACG GAGGGCCAGU GCGUGCCGAA ACCCGCACCA AACUUUAAGA UCGCCAUCUG 6600
 GAGGGUGGCG GCCUCAGAGU ACGCGGAGGU GACGCAGCAC GGGUCAUACC ACUACAUAAC 6660
 AGGACUUACC ACUGAUAAAU UGAAAGUUC UUGCCAACUA CCUUCUCCAG AGUUCUUUUC 6720
 CUGGGUGGAC GGAGUGCAGA UCCAUAAGGU UGCCCCCAUA CCGAAGCCGU UUUUUCGGGA 6780
 UGAGGUCUCG UUCUGCGUUG GGCUUAAUUC AUUUGUCGUC GGGUCUCAGC UCCCUGGCGA 6840
 UCCUGAACCU GACACAGACG UAUUGACGUC CAUGCUAACA GACCAUCCC AUAUCACGGC 6900
 GGAGACUGCA GCGCGGCGUU UGGCACGGGG GUCACCCCCG UCCGAGGCAA GCUCCUCAGC 6960
 GAGCCAGCUA UCGGCACCAU CGCUGCGAGC CACCUGCACC ACCCACGGCA AGGCCUAUGA 7020
 UGUGGACAUG GUGGAUGCCA ACCUGUUCAU GGGGGGCGAU GUGACCCGGA UAGAGUCUGA 7080
 GUCCAAAGUG GUCGUUCUGG ACUCUCUCGA CCCAUGGUC GAAGAAAGGA GCGACCUUGA 7140
 GCCUUCGAUA CCAUCGGAAU AUAUGCUCUCC CAAGAAGAGA UUCCCACCAG CCUUACCGGC 7200
 UUGGGCACGG CCUGAUUACA ACCCACCGCU UGUGGAAUCG UGGAAGAGGC CAGAUUACCA 7260
 ACCGGCCACU GUUGCGGGCU GCGCUCUCCC CCCCCUAAG AAAACCCCGA CGCCUCCCCC 7320
 AAGGAGACGC CGGACAGUGG GUCUGAGUGA GAGCUCCAUA GCAGAUGCCC UACAACAGCU 7380
 GGCCAUAAG UCCUUUGGCC AGCCCCCCCC AAGCGGCGAU UCAGGCCUUU CCACGGGGGC 7440
 GGACGCAGCC GAUUCGGGCA GUCGGACGCC CCCCAGAG UUGGCCUUU CGGAGACAGG 7500
 UUCCAUCUCC UCCAUGCCCC CUCUCGAGGG GGAGCCUGGA GAUCCAGACU UGGAGCCUGA 7560
 GCAGGUAGAG CUUCAACCUC CCCCCAGGG GGGGGUGGUA ACCCCCGGCU CAGGCUCGGG 7620
 GUCUUGGUCU ACUUGCUCUG AGGAGGACGA CUCCGUCGUG UGCUGCUCCA UGUCAUACUC 7680
 CUGGACCGGG GCUCUAAUAA CUCCUUGUAG CCCCAGAG GAAAAGUUGC CAAUUGGCC 7740
 CUUGAGCAAC UCCCUGUUGC GAUAUCACAA CAAGGUGUAC UGUACCACAU CAAAGAGCGC 7800
 CUCAUUAAGG GCUAAAAAGG UAACUUUUGA UAGGAUGCAA GCGCUCGACG CUCAUUAUGA 7860
 CUCAGUCUUG AAGGACAUUA AGCUAGCGGC CUCCAAGGUC ACCGCAAGGC UUCUCACUUU 7920
 AGAGGAGGCC UGCCAGUUA CUCCACCCA CUCUGCAAGA UCCAAGUAUG GGUUUGGGGC 7980
 UAAGGAGGUC CGCAGCUUGU CCGGGAGAGC CGUUAACCAC AUCAAGUCCG UGUGGAAGGA 8040
 CCUCCUGGAA GACACACAAA CACCAAUUCC UACAACCAUC AUGGCCAAA AUGAGGUGUU 8100

CUGCGUGGAC CCCACCAAGG GGGGUAAGAA AGCAGCUCGC CUUAUCGUUU ACCCUGACCU 8160
 CGGCGUCAGG GUCUGCGAGA AAAUGGCCC UUAUGAUAUC ACACAAAAGC UUCCUCAGGC 8220
 GGUGAUGGGG GCUUCUUAUG GAUUCCAGUA CUCCCCCGCU CAGCGGGUGG AGUUUCUCUU 8280
 GAAGGCAUGG GCGGAAAAGA AAGACCCUUA GGGUUUUUCG UAUGAUACCC GAUGC UUUGA 8340
 CUCAACCGUC ACUGAGAGAG ACAUCAGGAC UGAGGAGUCC AUUAUUCGGG CUUGUCCUU 8400
 GCCCAGGAG GCCACACUG CCAUACACUC ACUGACUGAG AGACUUUACG UGGGAGGGCC 8460
 CAUGUUAAC AGCAAGGGCC AGACCUGCGG GUACAGGCGU UGCCGCGCCA GCGGGUGCU 8520
 UACCACUAGC AUGGGGAACA CCAUCACAUG CUAUGUGAAA GCCUUAGCGG CCUGUAAGGC 8580
 UGCAGGGAUA AUUGCGCCA CAAUGCUGGU AUGCGGCGAU GACUUGGUUG UCAUCUCAGA 8640
 GAGCCAGGGG ACCGAGGAGG ACGAGCGGAA CCUGAGAGCC UUCACGGAGG CUAUGACCAG 8700
 GUUUUCUGCC CCUCCUGGUG ACCCCCCCAG ACCGGAUUAU GACCUGGAGC UGAUAACAUC 8760
 UUGCUCCUCA AAUGUGUCUG UGGCGUUGGG CCCACAAGGC CGCCGCAGAU ACUACCUGAC 8820
 CAGAGACCCU ACCACUCCAA UCGCCCGGGC UGCCUGGGAA ACAGUUAGAC ACUCCCCUGU 8880
 CAAUUAUGG CUAGGAAACA UCAUCCAGUA CGCCCCAACC AUAUGGGCUC GCAUGGUCCU 8940
 GAUGACACAC UUCUUCUCCA UUCUCAUGGC CCAAGAUACU CUGGACCAGA ACCUCAACUU 9000
 UGAGAUGUAC GGAGCGGUGU ACUCCGUGAG UCCCUUGGAC CUCCCAGCCA UAAUUGAAAG 9060
 GUUACACGGG CUUGACGCUU UCUCUCUGCA CACAUACACU CCCCACGAAC UGACACGGGU 9120
 GGCUCAGCC CUCAGAAAAC UUGGGGCGCC ACCCCUCAGA GCGUGGAAGA GCCGGGCACG 9180
 UGCAGUCAGG GCGUCCCUCA UCUCGUGG GGGGAGAGCG GCCGUUUGCG GCCGAUAUCU 9240
 CUUCAACUGG GCGGUGAAGA CCAAGCUCAA ACUCACUCCA UUGCCGGAAG CGCGCCUCCU 9300
 GGAUUUAUCC AGCUGGUUCA CUGUCGGCGC CGGCGGGGGC GACAUUUUUC ACAGCGUGUC 9360
 GCGUGCCGA CCCCUCUUAU UACUCCUUGG CCUACUCCUA CUUUUUGUAG GGGUAGGCCU 9420
 UUCCUACUC CCCGCUCCGU AGAGCGGCAC ACAUUAGCUA CACUCCAUA CUAACUGUCC 9480
 CUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU 9540
 UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU 9589

Sequence ID No.2
Sequence Length: 9,589
Sequence Type: nucleic acid
Strandedness: single
Topology: linear
Molecule Type: cDNA to genomic RNA
Method for Determination of Feature: E

```
ACCCGCCCCCT AATAGGGGCG ACACTCCGCC ATGAACCACT CCCCTGTGAG GAACTACTGT 60
CTTCACGCAG AAAGCGTCTA GCCATGGCGT TAGTATGAGT GTCGTACAGC CTCCAGGCC 120
CCCCCTCCCG GGAGAGCCAT AGTGGTCTGC GGAACCGGTG AGTACACCGG AATTGCCGGG 180
AAGACTGGGT CCTTTCTTGG ATAAACCCAC TCTATGCCCG GTCATTTGGG CGTGCCCCCG 240
CAAGACTGCT AGCCGAGTAG CGTTGGGTTG CGAAAGGCCT TGTGGTACTG CCTGATAGGG 300
TGCTTGCGAG TGCCCCGGGA GGTCTCGTAG ACCGTGCACC ATGAGCACA AATCCTAAACC 360
TCAAAGAAAA ACCAAAAGAA ACACCAACCG TCGCCACAAA GACGTTAAGT TTCCGGGCGG 420
CGGCCAGATC GTTGGCGGAG TATACTTGTT GCCGCGCAGG GGCCCCAGGT TGGGTGTGCG 480
CGCGACAAGG AAGACTTCGG AGCGGTCCCA GCCACGTGGA AGGCGCCAGC CCATCCCTAA 540
GGATCGGCGC TCCACTGGCA AATCCTGGGG AAAACCAGGA TACCCCTGGC CCCTATACGG 600
GAATGAGGGA CTCGGCTGGG CAGGATGGCT CCTGTCCCCC CGAGGTTCCC GTCCCTCTTG 660
GGGCCCCAAT GACCCCCGGC ATAGGTCCCG CAACGTGGGT AAGGTCATCG ATACCCTAAC 720
GTGCGGCTTT GCCGACCTCA TGGGGTACAT CCCTGTCGTA GCGGCCCCGC TCGGCGGCGT 780
CGCCAGAGCT CTCGCGCATG GCGTGAGAGT CCTGGAGGAC GGGGTTAATT TTGCAACAGG 840
GAACTTACCC GGTTGCTCCT TTTCTATCTT CTTGCTGGCC CTGCTGTCCT GCATCACCAC 900
CCCGGTCTCC GCTGCCGAAG TGAAGAACAT CAGTACCGGC TACATGGTGA CCAACGACTG 960
CACCAATGAT AGCATTACCT GGCAACTCCA GGCTGCTGTC CTCCACGTCC CCGGGTGCGT 1020
CCCGTGCGAG AAAGTGGGGA ATACATCTCG GTGCTGGATA CCGGTCTCAC CGAATGTGGC 1080
CGTGCAAGCAG CCCGGCGCCC TCACGCAGGG CTTACGGACG CACATTGACA TGGTTGTGAT 1140
GTCCGCCACG CTCTGCTCCG CTCTTTACGT GGGGGACCTC TGCGGTGGGG TGATGCTTGC 1200
```

AGCCCAGATG TTCATTGTCT CGCCACAGCA CCACTGGTTT GTGCAAGACT GCAATTGCTC 1260
 CATCTACCCT GGTACCATCA CTGGACACCG CATGGCGTGG GACATGATGA TGAAGTGGTC 1320
 GCCCACGGCT ACCATGATCC TGGCGTACGC GATGCGCGTC CCCGAGGTCA TCATAGACAT 1380
 CATTGGCGGG GCTCATTGGG GCGTCATGTT CGGCTTAGCC TACTTCTCTA TGCAGGGAGC 1440
 GTGGGCAAAA GTCGTTGTCA TTCTTTTGCT GGCCGCCGGG GTGGACGCGC AAACCCATAC 1500
 CGTTGGGGGT TCTACCGCGC ATAACGCCAG GACCCTCACC GGCATGTTCT CCCTTGGTGC 1560
 CAGGCAGAAA ATCCAGCTCA TCAACACCAA TGGCAGTTGG CACATCAACC GCACCGCCCT 1620
 GAACTGCAAT GACTCTTTGC ACACCGGCTT CCTCGCGTCA CTGTTCTACA CCCACAGCTT 1680
 CAACTCGTCA GGATGTCCCG AACGCATGTC CGCCTGCCGC AGTATCGAGG CCTTTCGGGT 1740
 GGGATGGGGC GCCTTACAAT ATGAGGACAA TGTACCAAT CCAGAGGATA TGAGACCGTA 1800
 TTGCTGGCAC TACCCACCAA GACAGTGTGG TGTAGTCTCC GCGAGCTCTG TGTGTGGCCC 1860
 AGTGTACTGT TTCACCCCCA GCCCAGTAGT AGTGGGTACG ACCGATAGAC TTGGAGCGCC 1920
 CACTTACACG TGGGGGGAGA ATGAGACAGA TGTCTTCTA TTGAACAGCA CTCGACCACC 1980
 GCAGGGGTCA TGGTTCGGCT GCACGTGGAT GAACTCCACT GGCTACACCA AGACTTGCGG 2040
 CGCACCAACC TGCCGCATTA GAGCTGACTT CAATGCCAGC ATGGACTTGT TGTGCCCCAC 2100
 GGAAGTGTTC AGGAAGCATC CTGATACCAC CTACATCAA TGTGGCTCTG GGCCCTGGCT 2160
 CAGCCAAGG TGCTGATCG ACTACCCCTA CAGGCTCTGG CATTACCCCT GCACAGTTAA 2220
 CTATACCATC TTCAAAATAA GGATGTAIGT GGGGGGGGTC GAGCACAGGC TCACGGCTGC 2280
 GTGCAATTC ACTCGTGGGG ATCGTTGCAA CTTGGAGGAC AGAGACAGAA GTCAACTGTC 2340
 TCCTTTGCTG CACTCCACCA CGGAGTGGGC CATTTTACCT TGCATTACT CGGACCTGCC 2400
 CGCCTTGTCG ACTGGTCTTC TCCACCICCA CCAAAACAIC GTGGACGTGC AATTCATGTA 2460
 TGGCCTATCA CCTGCTCTCA CAAAATACAT CGTCCGATGG GAGTGGGTAG TACTCTTATT 2520
 CCTGCTCTTA GCGGACGCCA GGGTTTGC GC CTGCTTATGG ATGCTCATCT TGTGGGCCA 2580
 GGCCGAAGCA GCACTAGAGA AGTTGGTCGT CTTGCACGCT GCGAGCGCAG CTAGCTGCAA 2640
 TGGCTTCCTA TACTTTGTCA TCTTTTTCGT GGCTGCTTGG TACATCAAGG GTCGGGTAGT 2700
 CCCCTTGGCT ACTTATCCC TCACTGGCCT ATGGTCCTTT GGCTACTGC TCCTAGCATT 2760
 GCCCCAACAG GCTTATGCTT ATGACGCATC TGTACATGGT CAGATAGGAG CAGCTCTGTT 2820
 GGTACTGATC ACTCTCTTCA CACTACCCCC CGGGTATAAG ACCCTTCTCA GCCGGTTTCT 2880
 GTGGTGGTTG TGCTATCTTC TGACCCTGGC GGAAGCTATG GTCCAGGAGT GGGCACCACC 2940

TATGCAGGTG CGCGGTGGCC GTGATGGGAT CATATGGGCC GTCGCCATAT TCTGCCCGGG 3000
 TGTGGTGTIT GACATAACCA AGTGGCTCTT GCGGGTGCTT GGGCCTGCTT ATCTCCTAAA 3060
 AGGTGCTTTG ACGCGTGTGC CGTACTTCGT CAGGGCTCAC GCTCTACTAA GGATGTGCAC 3120
 CATGGTAAGG CATCTCGCGG GGGGTAGGTA CGTCCAGATG GTGCTACTAG CCCTTGGCAG 3180
 GTGGA CTGGC ACTTACATCT ATGACCACCT CACCCCTATG TCGGATTGGG CTGCTAATGG 3240
 CCTGCGGGAC TTGGCGGTG CCGTGGAGCC TATCATCTTC AGTCCGATGG AGAAAAAGT 3300
 CATCGTCTGG GGAGCGGAGA CAGCTGCTTG CGGGGATATC TTACACGGAC TTCCCGTGTG 3360
 CGCCCGACTT GGCCGGGAGG TCCTCCTTGG CCCAGCTGAT GGCTATACCT CCAAGGGGTG 3420
 GAGTCTTCTC GCCCCCATCA CTGCTTATGC CCAGCAGACA CGCGGCCCTT TGGGCACCAT 3480
 AGTGGTGAGC ATGACGGGGC GCGACAAGAC AGAACAGGCC GGGGAGATTC AGGTCTGTG 3540
 CACGGTCACT CAGTCTTCC TCGGAACAAC CATCTCGGGG GTCTTATGGA CTGTCTACCA 3600
 TGGAGCTGGC AACAAGACTC TAGCCGGCTC ACGGGGTCCG GTCACACAGA TGTA CTCCAG 3660
 TGCTGAGGGG GACTTAGTGG GGTGGCCAG CCCCCCGGG ACCAAATCTT TGGAGCCGTG 3720
 CACGTGTGGA GCGGTGACC TATACCTGGT CACGCGAAAC GCTGATGTCA TCCCGGCTCG 3780
 AAGACGCGGG GACAAGCGAG GAGCGCTACT CTCCCCGAGA CCTCTTTCCA CCTTGAAGGG 3840
 GTCCTCGGGG GGCCCGGTGC TCTGCCCCAG AGGCCACGCT GTCGGGGTCT TCCGGGCAGC 3900
 CGTGTGCTCC CGGGGCGTGG CCAAGTCCAT AGATTTTATC CCCGTTGAGA CACTTGACAT 3960
 CGTCACTCGG TCCCCACCT TTAGTGACAA CAGCACACCA CCTGCTGTGC CCCAACTTA 4020
 TCAGGTGCGG TACTTACATG CCCCAGCTGG TAGTGGAAAG AGCACCAAAG TCCCTGTGCG 4080
 GTATGCCGCT CAGGGGTACA AAGTGCTAGT GCTTAATCCC TCGGTGGCTG CCACCCTGGG 4140
 GTTTGGGGCG TACTTGTTCA AGGCACATGG CATCAATCCC AACATTAGGA CTGGGGTCAG 4200
 GACTGTGACG ACCGGGGGCG CCATCACGTA CTCCACATAT GGCAAATTCC TCGCCGATGG 4260
 GGGCTGCGCA GCGGGCGCCT ATGACATCAT CATATGCGAT GAATGCCATG CCGTGGACTC 4320
 TACCACCATT CTCGGCATCG GAACAGTCTT CGATCAAGCA GAGACAGCCG GGGTCAGGCT 4380
 AACTGTACTG GCTACGGCTA CGCCCCCGG GTCAGTGACA ACCCCCCACC CCAACATAGA 4440
 GGAGGTGGCC CTCGGGCAGG AGGGTGAGAT CCCCTTCTAT GGGAGGGCGA TTCCCTGTG 4500
 ATACATCAAG GGAGGAAGAC ACTTGATCTT CTGCCACTCA AAGAAAAAGT GTGACGAGCT 4560
 CGCGGGCGCC CTTCGGGGTA TGGGCTTGAA CGCAGTGGCA TACTACAGAG GGCTGGACGT 4620
 CTCCGTAATA CCAACTCAGG GAGACGTAGT GGTGTCGCC ACCGACGCC TCATGACGGG 4680

GTTACTGGA GACTTTGACT CCGTGATCGA CTGCAACGTA GCGGTCACTC AAGTTGTAGA 4740
 CTTCAGCTTG GACCCACAT TCACCATAAC CACACAGACT GTCCCTCAAG ACGCTGTCTC 4800
 ACGTAGCCAG CGCCGGGGCC GCACGGGCAG GGGAAGACTG GGTATTTATA GGTATGTTTC 4860
 CACTGGTGAG CGAGCCTCAG GAATGTTTGA CAGTGTAGTG CTCTGCGAGT GCTACGATGC 4920
 AGGGGCCGCA TGGTATGAGC TCACACCAGC GGAGACCACC GTCAGGCTCA GAGCATATTT 4980
 CAACACACCT GGTTCGCTG TGTGCCAAGA CCATCTTGAG TTTGGGAGG CAGTTTTCAC 5040
 CGGCCTCACA CACATAGATG CCCACTTCCT TTCCCAAACA AAGCAATCGG GGGAAAATTT 5100
 CGCATACTTA ACAGCCTACC AGGCTACAGT GTGCGCTAGG GCCAAAGCCC CCCCCCGTC 5160
 CTGGGACGTC ATGTGGAAGT GTTTGACTCG ACTCAAGCCC AACTCGTGG GCCCCACACC 5220
 TCTCCTGTAC CGCTTGGGCT CTGTTACCAA CGAGGTCACC CTCACGCATC CTGTGACGAA 5280
 ATACATCGCC ACCTGCATGC AAGCCGACCT TGAGGTCATG ACCAGCACGT GGGTCTTAGC 5340
 TGGGGGGGTC TTGGCGGGCG TCGCCGCGTA CTGCCTGGCG ACCGGGTGTG TTTGCATCAT 5400
 CGGCCGCTTG CACGTTAACC AGCGAGCCGT CGTTGCACCG GACAAGGAGG TCCTCTATGA 5460
 GGCTTTTGAT GAGATGGAGG AATGTGCCTC TAGAGCGGCT CTCATTGAAG AGGGGCAGCG 5520
 GATAGCCGAG ATGCTGAAGT CCAAGATCCA AGGCTTATTG CAGCAAGCTT CCAAACAAGC 5580
 TCAAGACATA CAACCCGCTG TGCAGGCTTC TTGGCCCAAG GTAGAGCAAT TCTGGGCCAA 5640
 ACACATGTGG AACTTCATCA GCGGCATTCA ATACCTCGCA GGAATATCAA CACTGCCAGG 5700
 GAACCCGTGT GTAGCTTCCA TGATGGCATT CAGTGCCGCC CTCACCAGTC CGTTGTCAAC 5760
 TAGCACCCT ATCCTTCTCA ACATTTTGGG GGGCTGGCTA GCATCCCAA TTGCGCCTCC 5820
 CGCGGGGGCT ACCGGCTTCG TCGTCAGTGG CCTGGTGGGG GCTGCCGTAG GCAGCATAGG 5880
 CTGGGTAAG GTGCTGGTGG ACATCCTGGC AGGGTATGGT GCGGGCATTT CCGGGGCTCT 5940
 CGTCGCATTC AAGATCATGT CTGGCGAGAA GCCCTCCATG GAGGATGTTG TCAACCTGCT 6000
 GCCTGGAATT CTGTCTCCGG GTGCCCTGGT GGTGGGAGTC ATCTGCGCGG CCATCCTGCG 6060
 CCGACACGTG GGACCGGGGG AAGGCGCTGT CCAATGGATG AATAGGCTCA TTGCCTTTGC 6120
 TTCCAGAGGA AACCACGTG CCCCCACCCA CTACGTGACG GAGTCGGATG CGTCGCAGCG 6180
 TGTGACCCAA CTAATTGGCT CCCTTACCAT AACCAGCCTG CTCAGGAGAC TCCACAACCTG 6240
 GATTACTGAA GACTGCCCCA TCCCATGCAG CGGCTCGTGG CTCCGCGATG TGTGGGATTG 6300
 GGTTCGACC ATCCTAACAG ACTTTAAAA CTGGCTGACC TCCAAATTGT TCCCAAAGAT 6360
 GCCTGGTCTC CCCTTTATCT CTGTGCAAAA GGGGTACAAG GGCGTGTGGG CTGGCACTGG 6420

TATCATGACC ACACGGTGTC CTTGCGGCGC CAATATCTCT GGCAATGTCC GCCTGGGCTC 6480
CATGAGAATT ACGGGGCCCA AAACCTGCAT GAATATCTGG CAGGGGACCT TTCCCATCAA 6540
TTGTTACACG GAGGGCCAGT GCGTGCCGAA ACCCGCACCA AACTTTAAGA TCGCCATCTG 6600
GAGGGTGGCG GCCTCAGAGT ACGCGGAGGT GACGCAGCAC GGGTCATACC ACTACATAAC 6660
AGGACTTACC ACTGATAACT TGAAAGTTCC TTGCCAACTA CCTTCTCCAG AGTTCTTTTC 6720
CTGGGTGGAC GGAGTGCAGA TCCATAGGTT TGCCCCATA CCGAAGCCGT TTTTTCGGGA 6780
TGAGGTCTCG TTCTGCGTTG GGCTTAATTC ATTTGTCGTC GGGTCTCAGC TCCCTTGCGA 6840
TCCTGAACCT GACACAGACG TATTGACGTC CATGCTAACA GACCCATCCC ATATCACGGC 6900
GGAGACTGCA GCGCGGCGTT TGGCACGGGG GTCACCCCCG TCCGAGGCAA GCTCCTCAGC 6960
GAGCCAGCTA TCGGCACCAT CGCTGCGAGC CACCTGCACC ACCCACGGCA AGGCCTATGA 7020
TGTGGACATG GTGGATGCCA ACCTGTTTAT GGGGGGCGAT GTGACCCGGA TAGAGTCTGA 7080
GTCCAAAGTG GTCGTTCTGG ACTCTCTCGA CCAATGGTC GAAGAAAGGA GCGACCTTGA 7140
GCCTTCGATA CCATCGGAAT ATATGCTCCC CAAGAAGAGA TTCCCACCAG CTTACCGGC 7200
TTGGGCACGG CCTGATTACA ACCCACCGCT TGTGGAATCG TGGAAGAGGC CAGATTACCA 7260
ACCGGCCACT GTTGCGGGCT GCGCTCTCCC CCCCCCTAAG AAAACCCCGA CGCCTCCCCC 7320
AAGGAGACGC CGGACAGTGG GTCTGAGTGA GAGCTCCATA GCAGATGCCC TACAACAGCT 7380
GGCCATCAAG TCCTTTGGCC AGCCCCCCCC AAGCGGCGAT TCAGGCCTTT CCACGGGGGC 7440
GGACGCAGCC GATTCCGGCA GTCGGACGCC CCCCAGTGAG TTGGCCCTTT CGGAGACAGG 7500
TTCCATCTCC TCCATGCCCC CTCTCGAGGG GGAGCCTGGA GATCCAGACT TGGAGCCTGA 7560
GCAGGTAGAG CTTCAACCTC CCCCCAGGG GGGGGTGGTA ACCCCCGGCT CAGGCTCGGG 7620
GICITGGTCT ACTTGCTCCG AGGAGGACGA CTCGCTCGTG TGCTGCTCCA TGTCACTC 7680
CTGGACCGGG GCTCTAATAA CTCCTTGTAG CCCCAGAGAG GAAAAGTTGC CAATTGGCCC 7740
CTTGAGCAAC TCCCTGTTGC GATATCACAA CAAGGTGTAC TGTACCACAT CAAAGAGCGC 7800
CTCATTAAAG GCTAAAAAGG TAACTTTTGA TAGGATGCAA GCGCTCGACG CTCATTATGA 7860
CTCAGTCTTG AAGGACATTA AGCTAGCGGC CTCCAAGGTC ACCGCAAGGC TTCTCACTTT 7920
AGAGGAGGCC TGCCAGTTAA CTCCACCCCA CTCTGCAAGA TCCAAGTATG GGTGTTGGGGC 7980
TAAGGAGGTC CGCAGCTTGT CCGGGAGAGC CGTTAACCAC ATCAAGTCCG TGTGGAAGGA 8040
CCTCCTGGAA GACACACAAA CACCAATTCC TACAACCATC ATGGCCAAAA ATGAGGTGTT 8100
CTGCGTGGAC CCCACCAAGG GGGGTAAGAA AGCAGCTCGC CTTATCGTTT ACCCTGACCT 8160

CGGCGTCAGG GTCTGCGAGA AAATGGCCCT TTATGATATC ACACAAAAGC TTCCTCAGGC 8220
 GGTGATGGGG GCTTCTTATG GATTCCAGTA CTCCCCCGCT CAGCGGGTGG AGTTTCTCTT 8280
 GAAGGCATGG GCGGAAAAGA AAGACCCTAT GGGTTTTTCG TATGATACCC GATGCTTTGA 8340
 CTCAACCGTC ACTGAGAGAG ACATCAGGAC TGAGGAGTCC ATATATCGGG CTGTTCCTT 8400
 GCCCAGAGGAG GCCCACAACG CCATACACTC ACTGACTGAG AGACTTTACG TGGGAGGGCC 8460
 CATGTTCAAC AGCAAGGGCC AGACCTGCGG GTACAGGCGT TGCCGCGCCA GCGGGGTGCT 8520
 TACCACTAGC ATGGGGAACA CCATCACATG CTATGTGAAA GCCTTAGCGG CCTGTAAGGC 8580
 TGCAGGGATA ATTGCGCCCA CAATGCTGGT ATGCGGCGAT GACTTGGTTG TCATCTCAGA 8640
 GAGCCAGGGG ACCGAGGAGG ACGAGCGGAA CCTGAGAGCC TTCACGGAGG CTATGACCAG 8700
 GTATTCTGCC CCTCCTGGTG ACCCCCCCAG ACCGGAATAT GACCTGGAGC TGATAACATC 8760
 TTGCTCCTCA AATGTGTCTG TGGCGTTGGG CCCACAAGGC CGCCGCAGAT ACTACCTGAC 8820
 CAGAGACCCCT ACCACITCAA TCGCCCGGGC TGCCTGGGAA ACAGTTAGAC ACTCCCCTGT 8880
 CAATTCATGG CTAGGAAACA TCATCCAGTA CGCCCCAACC ATATGGGCTC GCATGGTCCT 8940
 GATGACACAC TTCTTCTCCA TTCTCATGGC CCAAGATACT CTGGACCAGA ACCTCAACTT 9000
 TGAGATGTAC GGAGCGGTGT ACTCCGTGAG TCCCTTGGAC CTCCCAGCCA TAATTGAAAG 9060
 GTTACACGGG CTTGACGCTT TCTCTCTGCA CACATACACT CCCCACGAAC TGACACGGGT 9120
 GGCTTCAGCC CTCAGAAAAC TTGGGGCGCC ACCCCTCAGA GCGTGGAAGA GCCGGGCACG 9180
 TGCAGTCAGG GCGTCCCTCA TCTCCCGTGG GGGGAGAGCG GCCGTTTGGC GCCGATATCT 9240
 CTTCAACTGG GCGGTGAAGA CCAAGCTCAA ACTCACTCCA TTGCCGGAAG CGCGCCTCCT 9300
 GGATTTATCC AGCTGGTTCA CTGTCGGCGC CGGCGGGGGC GACATTTATC ACAGCGTGTC 9360
 GCGTGCCCGA CCCCCTTAT TACTCCTTGG CCTACTCCTA CTTTTGTAG GGGTAGGCCT 9420
 TTTCTACTC CCCGCTCGGT AGAGCGGCAC ACATTAGCTA CACTCCATAG CTAAGTGTCC 9480
 CTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 9540
 TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 9589

Sequence ID No.3
 Sequence Length: 3,970
 Sequence Type: nucleic acid
 Strandedness: single
 Topology: linear
 Molecule Type: cDNA to genomic RNA
 Method for Determination of Feature: E

```

GGCATTACCC CTGCACAGTT AACTATACCA TCTTCAAAT AAGGATGTAT GTGGGGGGGG 60
TCGAGCACAG GCTCACGGCT GCGTGCAATT TCACTCGTGG GGATCGTTGC AACTTGGAGG 120
ACAGAGACAG AAGTCAACTG TCTCCTTTGC TGCCTCCAC CACGGAGTGG GCCATTTTAC 180
CTTGCACTTA CTCGGACCTG CCCGCCTTGT CGACTGGTCT TCTCCACCTC CACCAAAACA 240
TCGTGGACGT GCAATTCATG TATGGCCTAT CACCTGCTCT CACAAAATAC ATCGTCCGAT 300
GGGAGTGGGT AGTACTCTTA TTCCTGCTCT TAGCGGACGC CAGGGTTTGC GCCTGCTTAT 360
GGATGCTCAT CTGTGTTGGG CAGGCCGAAG CAGCACTAGA GAAGTTGGTC GTCTTGACAG 420
CTGCGAGCGC AGCTAGCTGC AATGGCTTCC TATACTTGT CATCTTTTTC GTGGCTGCTT 480
GGTACATCAA GGGTCGGGTA GTCCCCTTGG CTACTTATTC CCTCACTGGC CTATGGTCCT 540
TTGGCCTACT GCTCCTAGCA TTGCCCAAC AGGCTTATGC TTATGACGCA TCTGTACATG 600
GTCAGATAGG AGCAGCTCTG TTGGTACTGA TCACTCTCTT TACACTCACC CCCGGGTATA 660
AGACCCTTCT CAGCCGGTTT CTGTGGTGGT TGTGCTATCT TCTGACCCTG GCGGAAGCTA 720
TGGTCCAGGA GTGGGCACCA CCTATGCAGG TCGCGGGTGG CCGTGATGGG ATCATATGGG 780
CCGTCGCCAT ATTCTGCCCG GGTGTGGTGT TTGACATAAC CAAGTGGCTC TTGGCGGTGC 840
TTGGGCCTGC TTATCTCCTA AAAGGTGCTT TGACGCGTGT GCCGTACTTC GTCAGGGGCTC 900
ACGCTCTACT AAGGATGTGC ACCATGGTAA GGCATCTCGC GGGGGGTAGG TACGTCCAGA 960
TGGTGCTACT AGCCCTTGGC AGGTGGACTG GCACTTACAT CTATGACCAC CTCACCCCTA 1020
TGTCCGATTG GGCTGCTAAT GGCCTGCGGG ACTTGGCGGT CGCCGTGGAG CCTATCATCT 1080
TCAGTCCGAT GGAGAAAAAA GTCATCGTCT GGGGAGCGGA GACAGCTGCT TGCGGGGATA 1140
TCTTACACGG ACTTCCCGTG TCCGCCCCGAC TTGGCCGGGA GGTCCCTCCTT GGCCCAGCTG 1200

```

ATGGCTATAC CTCCAAGGGG TGGAGTCTTC TCGCCCCCAT CACTGCTTAT GCCCAGCAGA 1260
 CACGCGGCCT TTTGGGCACC ATAGTGGTGA GCATGACGGG GCGCGACAAG ACAGAACAGG 1320
 CCGGGGAGAT TCAGGTCCTG TCCACGGTCA CTCAGTCCTT CCTCGGAACA ACCATCTCGG 1380
 GGGTCTTATG GACTGTCTAC CATGGAGCTG GCAACAAGAC TCTAGCCGGC TCACGGGGTC 1440
 CGGTACACAC GATGTACTCC AGTGCTGAGG GGGACTTAGT GGGGTGGCCC AGCCCCCCCC 1500
 GGACCAAATC TTTGGAGCCG TGCACGTGTG GAGCGGTCTG CCTATACCTG GTCACGCGAA 1560
 ACGCTGATGT CATCCCGGCT CGAAGACGCG GGGACAAGCG AGGAGCGCTA CTCTCCCCGA 1620
 GACCTCTTTC CACCTTGAAG GGGTCCTCGG GGGGCCCCGT GCTCTGCCCC AGAGGCCACG 1680
 CTGTCGGGGT CTTCCGGGCA GCCGTGTGCT CCCGGGGCGT GGCCAAGTCC ATAGATTTTA 1740
 TCCCCGTTGA GACACTTGAC ATCGTCACTC GGTCCCCCAC CTTTAGTGAC AACAGCACAC 1800
 CACCTGCTGT GCCCCAACT TATCAGGTCG GGTACTTACA TGCCCCGACT GGTAGTGGAA 1860
 AGAGCACCAA AGTCCCTGTC GCGTATGCCG CTCAGGGGTA CAAAGTGCTA GTGCTTAATC 1920
 CCTCGGTGGC TGCCACCCTG GGGTTTGGGG CGTACTTGTC CAAGGCACAT GGCATCAATC 1980
 CCAACATTAG GACTGGGGTC AGGACTGTGA CGACCGGGGC GCCATCACG TACTCCACAT 2040
 ATGGCAAATT CCTCGCCGAT GGGGGCTGCG CAGGCGGCGC CTATGACATC ATCATATGCG 2100
 ATGAATGCCA TGCCGTGGAC TCTACCACCA TTCTCGGCAT CGGAACAGTC CTCGATCAAG 2160
 CAGAGACAGC CGGGGTGAGG CTAAGTGTAC TGGCTACGGC TACGCCCCCC GGGTCAGTGA 2220
 CAACCCCCCA CCCCAACATA GAGGAGGTGG CCCTCGGGCA GGAGGGTGAG ATCCCCTTCT 2280
 ATGGGAGGGC GATTCCCCTG TCATACATCA AGGGAGGAAG ACACTTGATC TTCTGCCACT 2340
 CAAAGAAAAA GTGTGACGAG CTCGCGGCGG CCCTTCGGGG TATGGGCTTG AACGCAGTGG 2400
 CATACTACAG AGGGCTGGAC GTCTCCGTAA TACCAACTCA GGGAGACGTA GTGGTCGTCTG 2460
 CCACCGACGC CCTCATGACG GGGTTTACTG GAGACTTTGA CTCCGTGATC GACTGCAACG 2520
 TAGCGGTAC TCAAGTTGTA GACTTCAGCT TGGACCCAC ATTACCCATA ACCACACAGA 2580
 CTGTCCCTCA AGACGCTGTC TCACGTAGCC AGCGCCGGGG CCGCACGGGC AGGGGAAGAC 2640
 TGGGTATTTA TAGGTATGTT TCCACTGGTG AGCGAGCCTC AGGAATGTTT GACAGTGTAG 2700
 TGCTCTGCGA GTGCTACGAT GCAGGGGCCG CATGGTATGA GCTCACACCA GCGGAGACCA 2760
 CCGTCAGGCT CAGAGCATAT TTCAACACAC CTGGTTTGCC TGTGTGCCAA GACCATCTTG 2820
 AGTTTTGGGA GCAGTTTTTC ACCGGCCTCA CACACATAGA TGCCCACTTC CTTTCCCAA 2880
 CAAAGCAATC GGGGGAAAAT TTCGCATACT TAACAGCCTA CCAGGCTACA GTGTGCGCTA 2940

GGGCCAAAGC CCCCCCCCCG TCCTGGGACG TCATGTGGAA GTGTTTGA CTGACTCAAGC 3000
CCACACTCGT GGGCCCCACA CCTCTCCTGT ACCGCTTGGG CTCTGTTACC AACGAGGTCA 3060
CCCTCACGCA TCCTGTGACG AAATACATCG CCACCTGCAT GCAAGCCGAC CTTGAGGTCA 3120
TGACCAGCAC GTGGGTCTTA GCTGGGGGGG TCTTGGCGGC CGTCGCCGCG TACTGCCTGG 3180
CGACCGGGTG TGTTTGATC ATCGGCCGCT TGCACGTTAA CCAGCGAGCC GTCGTTGCAC 3240
CGGACAAGGA GGTCTCTAT GAGGCTTTTG ATGAGATGGA GGAATGTGCC TCTAGAGCGG 3300
CTCTCATTGA AGAGGGGCAG CGGATAGCCG AGATGCTGAA GTCCAAGATC CAAGGCTTAT 3360
TGCAGCAAGC TTCCAAACAA GCTCAAGACA TACAACCCGC TGTGCAGGCT TCTTGGCCCA 3420
AGGTAGAGCA ATTCTGGGCC AAACACATGT GGAACCTCAT CAGCGGCATT CAATACCTCG 3480
CAGGACTATC AACACTGCCA GGAACCCCTG CTGTAGCTTC CATGATGGCA TTCAGTGCCG 3540
CCCTCACCAG TCCGTTGTCA ACTAGCACCA CTATCCTTCT CAACATTTTG GGGGGCTGGC 3600
TAGCATCCCA AATTGCGCCT CCCGCGGGGG CTACCGGCTT CGTCGTCAGT GGCCTGGTGG 3660
GGGCTGCCGT AGGCAGCATA GGCTTGGGTA AGGTGCTGGT GGACATCCTG GCAGGGTATG 3720
GTGCGGGCAT TTCGGGGGCT CTCGTGCGAT TCAAGATCAT GTCTGGCGAG AAGCCCTCCA 3780
TGGAGGATGT TGTC AACCTG CTGCCTGGAA TTCTGTCTCC GGGTGCCCTG GTGGTGGGAG 3840
TCATCTGCGC GGCCATCCTG CGCCGACACG TGGGACCGGG GGAAGGCGCT GTCCAATGGA 3900
TGAATAGGCT CATTGCCTTT GCTTCCAGAG GAAACCACGT CGCCCCACC CACTACGTGA 3960
CGGAGTCGGA 3970

Sequence ID No.4
 Sequence Length: 2.693
 Sequence Type: nucleic acid
 Strandedness: single
 Topology: linear
 Molecule Type: cDNA to genomic RNA
 Method for Determination of Feature: E

```

ATTCTGTCTC CGGGTGCCCT GGTGGTGGGA GTCATCTGCG CGGCCATCCT GCGCCGACAC   60
GTGGGACCGG GGAAGGGCG TGTCCAATGG ATGAATAGGC TCATTGCCTT TGCTTCCAGA  120
GGAAACCACG TCGCCCCAC CCACTACGTG ACGGAGTCGG ATGCGTCGCA GCGTGTGACC  180
CAACTACTTG GCTCCCTTAC CATAACCAGC CTGCTCAGGA GACTCCACAA CTGGATTACT  240
GAAGACTGCC CCAICCCAIG CAGCGGCTCG IGGCTCCGCG ATGTGTGGGA TTGGGTTTGC  300
ACCATCCTAA CAGACTTTAA AAAGTGGCTG ACCTCCAAAT TGTTCCTAAA GATGCCTGGT  360
CTCCCCTTTA TCTCTTGTC AAAGGGGTAC AAGGGCGTGT GGGCTGGCAC TGGTATCATG  420
ACCACACGGT GTCCTTGCGG CGCCAATATC TCTGGCAATG TCCGCCTGGG CTCCATGAGA  480
ATTACGGGGC CAAAAACCTG CATGAATATC TGGCAGGGGA CCTTTCCCAT CAATTGTTAC  540
ACGGAGGGCC AGTGCGTGCC GAAACCCGCA CCAAACCTTA AGATCGCCAT CTGGAGGGTG  600
GCGGCCTCAG AGTACGCGGA GGTGACGCAG CACGGGTCAT ACCACTACAT AACAGGACTT  660
ACCACTGATA ACTTGAAAGT TCCTTGCCAA CTACCTTCTC CAGAGTTCTT TTCCTGGGTG  720
GACGGAGTGC AGATCCATAG GTTTGCCCCC ATACCGAAGC CGTTTTTTTCG GGATGAGGTC  780
TCGTTCTGCG TTGGGCTTAA TTCATTTGTC GTCGGGTCTC AGCTCCCTTG CGATCCTGAA  840
CCTGACACAG ACGTATTGAC GTCCATGCTA ACAGACCCAT CCCATATCAC GCGGGAGACT  900
GCAGCGCGGC GTTTGGCACG GGGGTCACCC CCGTCCGAGG CAAGCTCCTC AGCGAGCCAG  960
CTATCGGCAC CATCGCTGCG AGCCACCTGC ACCACCCACG GCAAGGCCTA TGATGTGGAC 1020
ATGGTGGATG CCAACCTGTT CATGGGGGGC GATGTGACCC GGATAGAGTC TGAGTCCAAA 1080
GTGGTCGTTT TGGACTCTCT CGACCCAATG GTCGAAGAAA GGAGCGACCT TGAGCCTTCG 1140
ATACCATCGG AATATATGCT CCCCAGAAG AGATTCCAC CAGCCTTACC GGCTTGGGCA 1200

```


CGGCCTGATT ACAACCCACC GCTTGTGGAA TCGTGAAGA GGCCAGATTA CCAACCGGCC 1260
 ACTGTTGCGG GCTGCGCTCT CCCCCCCCCT AAGAAAACCC CGACGCCTCC CCCAAGGAGA 1320
 CGCCGGACAG TGGGTCTGAG TGAGAGCTCC ATAGCAGATG CCCTACAACA GCTGGCCATC 1380
 AAGTCCTTTG GCCAGCCCCC CCCAAGCGGC GATTCAAGGC TTTCCACGGG GGCGGACGCA 1440
 GCCGATTCCG GCAGTCGGAC GCCCCCGAT GAGTTGGCCC TTTCGGAGAC AGGTTCCATC 1500
 TCCTCCATGC CCCCTCTCGA GGGGGAGCCT GGAGATCCAG ACTTGGAGCC TGAGCAGGTA 1560
 GAGCTTCAAC CTCCCCCCCC GGGGGGGGTG GTAACCCCGG GCTCAGGCTC GGGGTCTTGG 1620
 TCTACTTGCT CCGAGGAGGA CGACTCCGTC GTGTGCTGCT CCATGTCATA CTCCTGGACC 1680
 GGGGCTCTAA TAACTCCTTG TAGCCCCGAA GAGGAAAAGT TGCCAATTGG CCCCTTGAGC 1740
 AACTCCCTGT TGCGATATCA CAACAAGGTG TACTGTACCA CATCAAAGAG CGCCTCATT 1800
 AGGGCTAAAA AGGTAACCTT TGATAGGATG CAAGCGCTCG ACGCTCATT 1860
 TGAAGGACA TTAAGCTAGC GGCCTCCAAG GTCACCGCAA GGCTTCTCAC TTTAGAGGAG 1920
 GCCTGCCAGT TAACTCCACC CCACTCTGCA AGATCCAAGT ATGGGTTTGG GGCTAAGGAG 1980
 GTCCGCAGCT TGTCCGGGAG AGCCGTTAAC CACATCAAGT CCGTGTGGAA GGACCTCCTG 2040
 GAAGACACAC AAACACCAAT TCCTACAACC ATCATGGCCA AAAATGAGGT GTTCTGCGTG 2100
 GACCCACCA AGGGGGGTAA GAAAGCAGCT CGCCTTATCG TTTACCCTGA CCTCGGCGTC 2160
 AGGGTCTGCG AGAAAATGGC CCTTTATGAT ATCACACAAA AGCTTCCTCA GGCGGTGATG 2220
 GGGGCTTCTT ATGGATTCCA GTACTCCCC GCTCAGCGGG TGGAGTTTCT CTTGAAGGCA 2280
 TGGGCGGAAA AGAAAGACCC TATGGGTTTT TCGTATGATA CCCGATGCTT TGAATCAACC 2340
 GTCACTGAGA GAGACATCAG GACTGAGGAG TCCATATATC GGGCTTGTTT CTTGCCCGAG 2400
 GAGGCCACA CTGCCATACA CTCCTGACT GAGAGACTTT ACGTGGGAGG GCCCATGTTC 2460
 AACAGCAAGG GCCAGACCTG CGGGTACAGG CGTTGCCGCG CCAGCGGGGT GCTTACCACT 2520
 AGCATGGGGA ACACCATCAC ATGCTATGTG AAAGCCTTAG CGGCCTGTAA GGCTGCAGGG 2580
 ATAATTGCGC CCACAATGCT GGTATGCGGC GATGACTTGG TTGTCATCTC AGAGAGCCAG 2640
 GGGACCGAGG AGGACGAGCG GAACCTGAGA GCCTTCACGG AGGCTATGAC CAG 2693

Sequence ID No.5
 Sequence Length: 3,033
 Sequence Type: amino acid
 Topology: linear
 Molecule Type: protein

Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn	Thr			
				5					10					15			
Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile			
				20					25					30			
Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	Leu	Gly			
				35					40					45			
Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	Pro	Arg	Gly			
				50					55					60			
Arg	Arg	Gln	Pro	Ile	Pro	Lys	Asp	Arg	Arg	Ser	Thr	Gly	Lys	Ser			
				65					70					75			
Trp	Gly	Lys	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly			
				80					85					90			
Leu	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro			
				95					100					105			
Ser	Trp	Gly	Pro	Asn	Asp	Pro	Arg	His	Arg	Ser	Arg	Asn	Val	Gly			
				110					115					120			
Lys	Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly			
				125					130					135			
Tyr	Ile	Pro	Val	Val	Gly	Ala	Pro	Leu	Gly	Gly	Val	Ala	Arg	Ala			
				140					145					150			
Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Phe	Ala			
				155					160					165			

Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala		
	170	175 180
Leu Leu Ser Cys Ile Thr Thr Pro Val Ser Ala Ala Glu Val Lys		
	185	190 195
Asn Ile Ser Thr Gly Tyr Met Val Thr Asn Asp Cys Thr Asn Asp		
	200	205 210
Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu His Val Pro Gly		
	215	220 225
Cys Val Pro Cys Glu Lys Val Gly Asn Thr Ser Arg Cys Trp Ile		
	230	235 240
Pro Val Ser Pro Asn Val Ala Val Gln Gln Pro Gly Ala Leu Thr		
	245	250 255
Gln Gly Leu Arg Thr His Ile Asp Met Val Val Met Ser Ala Thr		
	260	265 270
Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Gly Val Met		
	275	280 285
Leu Ala Ala Gln Met Phe Ile Val Ser Pro Gln His His Trp Phe		
	290	295 300
Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly Thr Ile Thr Gly		
	305	310 315
His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Ala		
	320	325 330
Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro Glu Val Ile Ile		
	335	340 345
Asp Ile Ile Gly Gly Ala His Trp Gly Val Met Phe Gly Leu Ala		
	350	355 360
Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val Val Val Ile Leu		
	365	370 375
Leu Leu Ala Ala Gly Val Asp Ala Gln Thr His Thr Val Gly Gly		

380	385	390
Ser Thr Ala His Asn Ala Arg Thr Leu Thr Gly Met Phe Ser Leu		
395	400	405
Gly Ala Arg Gln Lys Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp		
410	415	420
His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu His Thr		
425	430	435
Gly Phe Leu Ala Ser Leu Phe Tyr Thr His Ser Phe Asn Ser Ser		
440	445	450
Gly Cys Pro Glu Arg Met Ser Ala Cys Arg Ser Ile Glu Ala Phe		
455	460	465
Arg Val Gly Trp Gly Ala Leu Gln Tyr Glu Asp Asn Val Thr Asn		
470	475	480
Pro Glu Asp Met Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg Gln		
485	490	495
Cys Gly Val Val Ser Ala Ser Ser Val Cys Gly Pro Val Tyr Cys		
500	505	510
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Leu Gly		
515	520	525
Ala Pro Thr Tyr Thr Trp Gly Glu Asn Glu Thr Asp Val Phe Leu		
530	535	540
Leu Asn Ser Thr Arg Pro Pro Gln Gly Ser Trp Phe Gly Cys Thr		
545	550	555
Trp Met Asn Ser Thr Gly Tyr Thr Lys Thr Cys Gly Ala Pro Pro		
560	565	570
Cys Arg Ile Arg Ala Asp Phe Asn Ala Ser Met Asp Leu Leu Cys		
575	580	585
Pro Thr Asp Cys Phe Arg Lys His Pro Asp Thr Thr Tyr Ile Lys		
590	595	600

Cys Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Leu Ile Asp Tyr		
605	610	615
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Tyr Thr Ile		
620	625	630
Phe Lys Ile Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Thr		
635	640	645
Ala Ala Cys Asn Phe Thr Arg Gly Asp Arg Cys Asn Leu Glu Asp		
650	655	660
Arg Asp Arg Ser Gln Leu Ser Pro Leu Leu His Ser Thr Thr Glu		
665	670	675
Trp Ala Ile Leu Pro Cys Thr Tyr Ser Asp Leu Pro Ala Leu Ser		
680	685	690
Thr Gly Leu Leu His Leu His Gln Asn Ile Val Asp Val Gln Phe		
695	700	705
Met Tyr Gly Leu Ser Pro Ala Leu Thr Lys Tyr Ile Val Arg Trp		
710	715	720
Glu Trp Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val		
725	730	735
Cys Ala Cys Leu Trp Met Leu Ile Leu Leu Gly Gln Ala Glu Ala		
740	745	750
Ala Leu Glu Lys Leu Val Val Leu His Ala Ala Ser Ala Ala Ser		
755	760	765
Cys Asn Gly Phe Leu Tyr Phe Val Ile Phe Phe Val Ala Ala Trp		
770	775	780
Tyr Ile Lys Gly Arg Val Val Pro Leu Ala Thr Tyr Ser Leu Thr		
785	790	795
Gly Leu Trp Ser Phe Gly Leu Leu Leu Leu Ala Leu Pro Gln Gln		
800	805	810
Ala Tyr Ala Tyr Asp Ala Ser Val His Gly Gln Ile Gly Ala Ala		

815	820	825
Leu Leu Val Leu Ile Thr Leu Phe Thr Leu Thr Pro Gly Tyr Lys		
830	835	840
Thr Leu Leu Ser Arg Phe Leu Trp Trp Leu Cys Tyr Leu Leu Thr		
845	850	855
Leu Ala Glu Ala Met Val Gln Glu Trp Ala Pro Pro Met Gln Val		
860	865	870
Arg Gly Gly Arg Asp Gly Ile Ile Trp Ala Val Ala Ile Phe Cys		
875	880	885
Pro Gly Val Val Phe Asp Ile Thr Lys Trp Leu Leu Ala Val Leu		
890	895	900
Gly Pro Ala Tyr Leu Leu Lys Gly Ala Leu Thr Arg Val Pro Tyr		
905	910	915
Phe Val Arg Ala His Ala Leu Leu Arg Met Cys Thr Met Val Arg		
920	925	930
His Leu Ala Gly Gly Arg Tyr Val Gln Met Val Leu Leu Ala Leu		
935	940	945
Gly Arg Trp Thr Gly Thr Tyr Ile Tyr Asp His Leu Thr Pro Met		
950	955	960
Ser Asp Trp Ala Ala Asn Gly Leu Arg Asp Leu Ala Val Ala Val		
965	970	975
Glu Pro Ile Ile Phe Ser Pro Met Glu Lys Lys Val Ile Val Trp		
980	985	990
Gly Ala Glu Thr Ala Ala Cys Gly Asp Ile Leu His Gly Leu Pro		
995	1000	1005
Val Ser Ala Arg Leu Gly Arg Glu Val Leu Leu Gly Pro Ala Asp		
1010	1015	1020
Gly Tyr Thr Ser Lys Gly Trp Ser Leu Leu Ala Pro Ile Thr Ala		
1025	1030	1035

Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Thr Ile Val Val Ser	1040	1045	1050
Met Thr Gly Arg Asp Lys Thr Glu Gln Ala Gly Glu Ile Glu Val	1055	1060	1065
Leu Ser Thr Val Thr Gln Ser Phe Leu Gly Thr Thr Ile Ser Gly	1070	1075	1080
Val Leu Trp Thr Val Tyr His Gly Ala Gly Asn Lys Thr Leu Ala	1085	1090	1095
Gly Ser Arg Gly Pro Val Thr Gln Met Tyr Ser Ser Ala Glu Gly	1100	1105	1110
Asp Leu Val Gly Trp Pro Ser Pro Pro Gly Thr Lys Ser Leu Glu	1115	1120	1125
Pro Cys Thr Cys Gly Ala Val Asp Leu Tyr Leu Val Thr Arg Asn	1130	1135	1140
Ala Asp Val Ile Pro Ala Arg Arg Arg Gly Asp Lys Arg Gly Ala	1145	1150	1155
Leu Leu Ser Pro Arg Pro Leu Ser Thr Leu Lys Gly Ser Ser Gly	1160	1165	1170
Gly Pro Val Leu Cys Pro Arg Gly His Ala Val Gly Val Phe Arg	1175	1180	1185
Ala Ala Val Cys Ser Arg Gly Val Ala Lys Ser Ile Asp Phe Ile	1190	1195	1200
Pro Val Glu Thr Leu Asp Ile Val Thr Arg Ser Pro Thr Phe Ser	1205	1210	1215
Asp Asn Ser Thr Pro Pro Ala Val Pro Gln Thr Tyr Gln Val Gln	1220	1225	1230
Tyr Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	1235	1240	1245
Val Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro			

1250	1255	1260
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Leu Ser Lys Ala		
1265	1270	1275
His Gly Ile Asn Pro Asn Ile Arg Thr Gly Val Arg Thr Val Thr		
1280	1285	1290
Thr Gly Ala Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala		
1295	1300	1305
Asp Gly Gly Cys Ala Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp		
1310	1315	1320
Glu Cys His Ala Val Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr		
1325	1330	1335
Val Leu Asp Gln Ala Glu Thr Ala Gly Val Arg Leu Thr Val Leu		
1340	1345	1350
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Thr Pro His Pro Asn		
1355	1360	1365
Ile Glu Glu Val Ala Leu Gly Gln Glu Gly Glu Ile Pro Phe Tyr		
1370	1375	1380
Gly Arg Ala Ile Pro Leu Ser Tyr Ile Lys Gly Gly Arg His Leu		
1385	1390	1395
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Ala		
1400	1405	1410
Leu Arg Gly Met Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu		
1415	1420	1425
Asp Val Ser Val Ile Pro Thr Gln Gly Asp Val Val Val Val Ala		
1430	1435	1440
Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val		
1445	1450	1455
Ile Asp Cys Asn Val Ala Val Thr Gln Val Val Asp Phe Ser Leu		
1460	1465	1470

Asp Pro Thr Phe Thr Ile Thr Thr Gln Thr Val Pro Gln Asp Ala	1475	1480	1485
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Leu	1490	1495	1500
Gly Ile Tyr Arg Tyr Val Ser Thr Gly Glu Arg Ala Ser Gly Met	1505	1510	1515
Phe Asp Ser Val Val Leu Cys Glu Cys Tyr Asp Ala Gly Ala Ala	1520	1525	1530
Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala	1535	1540	1545
Tyr Phe Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu	1550	1555	1560
Phe Trp Glu Ala Val Phe Thr Gly Leu Thr His Ile Asp Ala His	1565	1570	1575
Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Ala Tyr Leu	1580	1585	1590
Thr Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Lys Ala Pro Pro	1595	1600	1605
Pro Ser Trp Asp Val Met Trp Lys Cys Leu Thr Arg Leu Lys Pro	1610	1615	1620
Trp Leu Val Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ser Val	1625	1630	1635
Thr Asn Glu Val Thr Leu Thr His Pro Val Thr Lys Tyr Ile Ala	1640	1645	1650
Thr Cys Met Gln Ala Asp Leu Glu Val Met Thr Ser Thr Trp Val	1655	1660	1665
Leu Ala Gly Gly Val Leu Ala Ala Val Ala Ala Tyr Cys Leu Ala	1670	1675	1680
Thr Gly Cys Val Cys Ile Ile Gly Arg Leu His Val Asn Gln Arg			

1685	1690	1695
Ala Val Val Ala Pro Asp Lys Glu Val Leu Tyr Glu Ala Phe Asp		
1700	1705	1710
Glu Met Glu Glu Cys Ala Ser Arg Ala Ala Leu Ile Glu Glu Gly		
1715	1720	1725
Gln Arg Ile Ala Glu Met Leu Lys Ser Lys Ile Gln Gly Leu Leu		
1730	1735	1740
Gln Gln Ala Ser Lys Gln Ala Gln Asp Ile Gln Pro Ala Val Gln		
1745	1750	1755
Ala Ser Trp Pro Lys Val Glu Gln Phe Trp Ala Lys His Met Trp		
1760	1765	1770
Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu		
1775	1780	1785
Pro Gly Asn Pro Ala Val Ala Ser Met Met Ala Phe Ser Ala Ala		
1790	1795	1800
Leu Thr Ser Pro Leu Ser Thr Ser Thr Thr Ile Leu Leu Asn Ile		
1805	1810	1815
Leu Gly Gly Trp Leu Ala Ser Gln Ile Ala Pro Pro Ala Gly Ala		
1820	1825	1830
Thr Gly Phe Val Val Ser Gly Leu Val Gly Ala Ala Val Gly Ser		
1835	1840	1845
Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly		
1850	1855	1860
Ala Gly Ile Ser Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly		
1865	1870	1875
Glu Lys Pro Ser Met Glu Asp Val Val Asn Leu Leu Pro Gly Ile		
1880	1885	1890
Leu Ser Pro Gly Ala Leu Val Val Gly Val Ile Cys Ala Ala Ile		
1895	1900	1905

Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met		
1910	1915	1920
Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ala Pro		
1925	1930	1935
Thr His Tyr Val Thr Glu Ser Asp Ala Ser Gln Arg Val Thr Gln		
1940	1945	1950
Leu Leu Gly Ser Leu Thr Ile Thr Ser Leu Leu Arg Arg Leu His		
1955	1960	1965
Asn Trp Ile Thr Glu Asp Cys Pro Ile Pro Cys Ser Gly Ser Trp		
1970	1975	1980
Leu Arg Asp Val Trp Asp Trp Val Cys Thr Ile Leu Thr Asp Phe		
1985	1990	1995
Lys Asn Trp Leu Thr Ser Lys Leu Phe Pro Lys Met Pro Gly Leu		
2000	2005	2010
Pro Phe Ile Ser Cys Gln Lys Gly Tyr Lys Gly Val Trp Ala Gly		
2015	2020	2025
Thr Gly Ile Met Thr Thr Arg Cys Pro Cys Gly Ala Asn Ile Ser		
2030	2035	2040
Gly Asn Val Arg Leu Gly Ser Met Arg Ile Thr Gly Pro Lys Thr		
2045	2050	2055
Cys Met Asn Ile Trp Gln Gly Thr Phe Pro Ile Asn Cys Tyr Thr		
2060	2065	2070
Glu Gly Gln Cys Val Pro Lys Pro Ala Pro Asn Phe Lys Ile Ala		
2075	2080	2085
Ile Trp Arg Val Ala Ala Ser Glu Tyr Ala Glu Val Thr Gln His		
2090	2095	2100
Gly Ser Tyr His Tyr Ile Thr Gly Leu Thr Thr Asp Asn Leu Lys		
2105	2110	2115
Val Pro Cys Gln Leu Pro Ser Pro Glu Phe Phe Ser Trp Val Asp		

Arg Ser Leu Ser Gly Arg Ala Val Asn His Ile Lys Ser Val Trp			
	2555	2560	2565
Lys Asp Leu Leu Glu Asp Thr Gln Thr Pro Ile Pro Thr Thr Ile			
	2570	2575	2580
Met Ala Lys Asn Glu Val Phe Cys Val Asp Pro Thr Lys Gly Gly			
	2585	2590	2595
Lys Lys Ala Ala Arg Leu Ile Val Tyr Pro Asp Leu Gly Val Arg			
	2600	2605	2610
Val Cys Glu Lys Met Ala Leu Tyr Asp Ile Thr Gln Lys Leu Pro			
	2615	2620	2625
Gln Ala Val Met Gly Ala Ser Tyr Gly Phe Gln Tyr Ser Pro Ala			
	2630	2635	2640
Gln Arg Val Glu Phe Leu Leu Lys Ala Trp Ala Glu Lys Lys Asp			
	2645	2650	2655
Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val			
	2660	2665	2670
Thr Glu Arg Asp Ile Arg Thr Glu Glu Ser Ile Tyr Arg Ala Cys			
	2675	2680	2685
Ser Leu Pro Glu Glu Ala His Thr Ala Ile His Ser Leu Thr Glu			
	2690	2695	2700
Arg Leu Tyr Val Gly Gly Pro Met Phe Asn Ser Lys Gly Gln Thr			
	2705	2710	2715
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser			
	2720	2725	2730
Met Gly Asn Thr Ile Thr Cys Tyr Val Lys Ala Leu Ala Ala Cys			
	2735	2740	2745
Lys Ala Ala Gly Ile Ile Ala Pro Thr Met Leu Val Cys Gly Asp			
	2750	2755	2760

Asp Leu Val Val Ile Ser Glu Ser Gln Gly Thr Glu Glu Asp Glu	2765	2770	2775
Arg Asn Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala	2780	2785	2790
Pro Pro Gly Asp Pro Pro Arg Pro Glu Tyr Asp Leu Glu Leu Ile	2795	2800	2805
Thr Ser Cys Ser Ser Asn Val Ser Val Ala Leu Gly Pro Gln Gly	2810	2815	2820
Arg Arg Arg Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Ile Ala	2825	2830	2835
Arg Ala Ala Trp Glu Thr Val Arg His Ser Pro Val Asn Ser Trp	2840	2845	2850
Leu Gly Asn Ile Ile Gln Tyr Ala Pro Thr Ile Trp Ala Arg Met	2855	2860	2865
Val Leu Met Thr His Phe Phe Ser Ile Leu Met Ala Gln Asp Thr	2870	2875	2880
Leu Asp Gln Asn Leu Asn Phe Glu Met Tyr Gly Ala Val Tyr Ser	2885	2890	2895
Val Ser Pro Leu Asp Leu Pro Ala Ile Ile Glu Arg Leu His Gly	2900	2905	2910
Leu Asp Ala Phe Ser Leu His Thr Tyr Thr Pro His Glu Leu Thr	2915	2920	2925
Arg Val Ala Ser Ala Leu Arg Lys Leu Gly Ala Pro Pro Leu Arg	2930	2935	2940
Ala Trp Lys Ser Arg Ala Arg Ala Val Arg Ala Ser Leu Ile Ser	2945	2950	2955
Arg Gly Gly Arg Ala Ala Val Cys Gly Arg Tyr Leu Phe Asn Trp	2960	2965	2970

Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Leu Pro Glu Ala Arg		
	2975	2980 2985
Leu Leu Asp Leu Ser Ser Trp Phe Thr Val Gly Ala Gly Gly Gly		
	2990	2995 3000
Asp Ile Tyr His Ser Val Ser Arg Ala Arg Pro Arg Leu Leu Leu		
	3005	3010 3015
Leu Gly Leu Leu Leu Leu Phe Val Gly Val Gly Leu Phe Leu Leu		
	3020	3025 3030
Pro Ala Arg		
	3033	

Sequence ID No. **6**

Sequence Length: 9,511

Sequence Type: nucleic acid

Strandedness: single

Topology: linear

Molecule Type: genomic RNA

Method for Determination of Feature: E

```

GCCCCGCCCC UGAUGGGGGC GACACUCCGC CAUGAAUCAC UCCCCUGUGA GGAACUACUG   60
UCUUCACGCA GAAAGCGUCU AGCCAUGGCG UUAGUAUGAG UGUCGUACAG CCUCCAGGCC  120
CCCCCUCCCC GGGAGAGCCA UAGUGGUCUG CGGAACCGGU GAGUACACCG GAAUUACCGG  180
AAAGACUGGG UCCUUUCUUG GAUAAACCCA CUCUAUGUCC GGUCAUUUGG GCACGCCCCC  240
GCAAGACUGC UAGCCGAGUA GCGUUGGGUU GCGAAAGGCC UUGUGGUACU GCCUGAUAGG  300
GURCUUGCGA GUGCCCCGGG AGGUCUCGUA GACCGUGCAU CAUGAGCACA AAUCCUAAAC  360
CUCAAAGAAA AACCAAAAGA AACACAAACC GCCGCCACA GGACGUUAAG UUCCCGGGUG  420
GCGGUCAGAU CGUUGGCGGA GUUUACUUGC UGCCGCGCAG GGGCCCCAGG UUGGGUGUGC  480
GCGCGACAAG GAAGACUUCY GAGCGAUCCC AGCCGCGUGG ACGACGCCAG CCCAUCCCGA  540
AAGAUCGGCG CUCCACCGGC AAGUCCUGGG GAAAGCCAGG AUAUCCUUGG CCCCUGUACG  600
GAAACGAGGG UUGCGGCUGG GCGGGUUGGC UCCUGUCCCC CCGCGGGUCU CGUCCUACUU  660
GGGGCCCCAC CGACCCCCGG CAUAGAUAC GCAAUUUGGG CAGAGUCAUC GAUACCAUUA  720
CGUGUGGUUU UGCCGACCUC AUGGGGUACA UCCUGUCGU UGGCGCCCCG GUYGGAGGCG  780
UCGCCAGAGC UCUGGCACAC GGUGUUAGGG UCCUGGAGGA CGGGAUAAAU UACGCAACAG  840
GGAAUUUACC CGGUUGCUCU UUUUCUAUCU UUUUGCUUGC UCUUCUGUCA UGCGUCACAR  900
UGCCAGUGUC UGCAGUGGAA GUCAGGAACA UYAGUUCUAG CUACUACGCC ACUAAUGAUU  960
GCUAAACAA CAGCAUCACC UGGCAGCUA CUGACGCAGU UCUCCAUCUU CCUGGAUGCG 1020
UCCCAUGUGA GAAYGAUAY GGCACCUUGC RUUGCUGGAU ACAAGUACA CCCRACGUGG 1080
CUGUGAAACA CCGCGGUGCG CUCACUCGUA GCCUGCGAAC ACACGUCGAC AUGAUCGUAA 1140

```

UGGCAGCUAC GGGCUGCUCG GCCUUGUAUG UGGGAGAUGU GUGCGGGGCC GUGAUGAUYC 1200
 UAUCGCAGGC UUUCAUGGUA UCACCACAAC GCCACAACUU CACCCAAGAG UGCAACUGUU 1260
 CCAUCUACCA AGGUCACAUC ACCGGCCAUC GCAUGGCAUG GGACAUGAUG CURARCUGGU 1320
 CUCCAACUCU URCCAUGAUC CUCGCCUACG CYGCUCGYGU UCCCGARCUG GUCCUCGAAA 1380
 UYAUUUUCGG CGGCCAUUGG GGUGUGGYGU UYGGCUUGGS CUAUUUCUCC AUGCARGGAG 1440
 CGUGGGCCAA AGUCRUYGCC AUCCUCCUUC UUGUUGCGGG AGUGGAUGCA WCCACCUAUU 1500
 CCASC GG YCA GSAAGCGGGU CGURCCGYCK HKGGGWUCKC URGCCUCUUU AMUACUGGUG 1560
 CCAAGCAGAA CCUCYAUUUR AUCAACACCA AUGGCAGCUG GCACAUAAAC CGGACUGCCC 1620
 UCAAUUGCAA UGACAGCYUA SAGACGGGUU UCHUCGCUUC CYUGKUUUAC WMCCRCARGU 1680
 UCAACAGCUC UGGCUGCCCC GAGCGCUUGU CUUCCUGCCG CGGGCUGGAC GAYUUYCGCA 1740
 UCGGCUGGGG AACCUUGGAA UACGAAACCA ACGUCACCAA CGAUGRGGAC AUGAGGCCGU 1800
 ACUGCUGGCA UUACCCCCCG AGGCCUUGCG GCAUCGUCCG GGCUAGGACG GUUUGCGGAC 1860
 CGGUCUAUUG YUUCACCCCU AGCCCUUGUUG UCGUGGGCAC CACUGACAAG CAGGGCGUAC 1920
 CCACCUACAC CUGGGGRGAA AACGAGACCG AUGUCUCCU GCRAAAUAGC ACAAGACCCC 1980
 CGCGAGGAGC UUGGUUCGGC UGCACYUGGA UGAACGGGAC UGGGUUCACU AAGACAUGCG 2040
 GUGCACCACC UUGCCGCAUU AGGAAAGACU ACAACAGCAC UCUCGAUUUA UUGUGCCCCA 2100
 CAGACUGUUU UAGGAAGCAC CCAGAUGCUA CCUAUCUUAU GUGUGGAGCA GGGCCUUGGU 2160
 UAACUCCCAG GUGCCUGGUA GACUACCCUU AUAGRYUGUG GCAUUAUCCG UGCACUGUAA 2220
 ACUUCACCAU CUUYAAGGCG CGGAUGUAUG UAGGAGGGGU GGAGCAUCGA UUCUCCGCAG 2280
 CAUGCAACUU CACGCGCGGA GAUCGCUGCA GACUGGAAGA UAGGGAUAGG GGYCAGCAGA 2340
 GUCCACUGCU GCAUUCCACU ACUGAGUGGG CGGUGYUCC AUGCUCCUUC UCUGACCUAC 2400
 CAGCACUAUC CACUGGCCUA UUGCACCUC ACCAAAACAU CGUGGACGUG CAGUACCUYU 2460
 ACGGACUUUC UCCGGCUCUG ACAAGAUACA UCGUGAAGUG GGAGUGGGUG AUCCUCCUUU 2520
 UCUUGUUGUU GGCAGACGCC AGGRUCUGUG CAUGCCUUUG GAUGCUCAWC AUACUGGGCC 2580
 AAGCCGAAGC GGCUCUUGAG AAGCUCAUCA UCUUGCACUC CGCUAGYGCU GCUAGUGCCA 2640
 AUGGUCCGCU GUGGUUUUUC AUCUUCUUUA CAGCGGCCUG GUACUUAAG GGCAGGGUGG 2700
 UCCCCGUGGC CACGUACUCU GUBCUCGGCU URUGGUCCUU CCUCCUCCUA GUCCUGGCYU 2760
 UACCACAGCA GGCUUAUGCC UUGGACGCUG CUGAACAAAGG GGAACUGGGG CUGGCCAUAU 2820
 UAGUAAUUAU AUCCAUCUUU ACUCUUACCC CAGCAUACAA GAUCCUCCUG AGCCGUUCAG 2880

UGUGGUGGCU GUCCUACAUG CUGGUCUUGG CCGAGGCCCA GAUUCAGCAA UGGGUUCCCC 2940
 CCCUGGAGGU CCGAGGGGGG CGUGACGGGA UCAUCUGGGU GGCUGUCAUU CUACACCCAC 3000
 GCCUUGUGUU UGAGGUCACG AAAUGGUUGU UAGCAAUCCU GGGGCCUGCC UACCUCUUUA 3060
 RAGCGUCUCU GCUACGGUAU CCGUACUUUG UGAGGGGCCA CGCUUUGCUA CGAGUGUGUA 3120
 CCCUGGUGAA ACACCUCGCR GGGGCUAGGU ACAUCCAGAU GCUGUURAUC ACCAUAGGCA 3180
 GAUGGACCGG CACUUACAUC UACGACCACC UCUCUUUUU AUCAACUUGG GCGGCCCAGG 3240
 GUUURCGGGA CCUGGCAAUC GCCGUGGAGC CUGUGGUGUU CAGCCCAAUG GAGAAGAAGG 3300
 UCAUUGUGUG GGGGGCUGAG ACAGUGGCGU GUGGAGACAU CCUGCAUGGC CUCCCGGUCU 3360
 CCGCGAGGCU AGGUAGGGAR GUUCUGCUCG GCCUGCCGA CGGCUACACC UCCAAGGGGU 3420
 GGAAKUCCU AGCUCCCAUU ACUGCUUACA CUCAGCAAAC UCGUGGUCUC CUGGGUGCUA 3480
 UCGUGGUCAG CCUAACGGGC CGCGACAAA AUGAGCAGGC UGGGCAGGUC CAGGUUCUGU 3540
 CCUCCGUCAC ACAAACUUUC UUGGGGACAU CCAUUUCGGG CGUCCUCUGG ACAGUAUAUC 3600
 ACGGGGCUGG UAAUAAGACC UUGGCCGGCC CCAAGGGACC AGUCACUCAG AUGUACACCA 3660
 GCGCAGAAGG GGACCUCGUG GGAUGGCCUA GUCCCCCGG GACUAAGUCA UUGGACCCCU 3720
 GUACCUGCGG GGCCGUAGAC CUCUACCUGG UCACCCGAAA CGCUGAUGUC AUUCCGGUCC 3780
 GGAGGAAAGA UGACCGACGG GGUGCAUUAC UCUCGCCAAG GCCCCUCUA ACCCUCAAAG 3840
 GAUCAUCCGG AGGGCCCGUG CUCUGCUCWA GGGGACACGC CGUGGGCUUG UUCAGAGCGG 3900
 CCGUGUGUGC CAGGGGUGUA GCCAAUCUA UUGACUUAU CCCCUGCAA UCACUCGAUR 3960
 UCGCCACAGG GACGCCCAGU UUCUCUGACA ACAGURCGCC GCCAGCUGUG CCCAGUCUU 4020
 ACCAGGUGGG UUACUUGCAC GCACCAACAG GCAGCGGAAA GAGCACCAAG GUCCCUGCCG 4080
 CGUAUGCCAG UCAGGGGUUU AAAGUACUCG UACUAAUCC CUCUGUCGG GCCACACUUG 4140
 GUUUUGGGGC CUACAUGUCC AAAGCCCACG GGAUCAACCC UAAUAUCAGA ACUGGAGUGC 4200
 GGACCGUUA CACCGGGGAC UCUAUCACUU ACUCCACUUA UGGCAAGUUU AUCGCAGAUG 4260
 GAGGCUUGUC AGCCGGUGCC UAUGACAUA UCAUAUGCGA CGAAUGCCAU UCAGUGGACG 4320
 CUACUACCAU CCUUGGCAUU GGAACAGUCC UUGACCAAGC UGAGACCGCA GGCGUCAGGC 4380
 UAGUGGUYUU GGCCACAGCC ACGCCUCCG GUACGGUGAC AACUCCCCAC AGUAAAUAG 4440
 AGGAGGUGGC CCUUGGUCAC GAGGGCGAGA UCCUUUUUA UGGCAAAGCU AUUCCCCUAG 4500
 CUUUAUCAA GGGGGGCAGA CACUUGAUCU UUUGCCAUUC AAAGAAGAAG UGCGACGAGC 4560
 UCGCAGCGGC CCUCCGGGGC AYGGGUGUA AUGCCGUUGC AUACUAUAGG GGUCUCGACG 4620

UCUCCGUUAU	ACCAACUCA	GGAGACGUGG	UGGUUGUCGC	CACUGAUGCC	CUA AUGACUG	4680
GGUACACCGG	CGACUUUGAC	UCYGUCAUCG	ACUGUAAUGU	UGCAGUCUCU	CAGAUUGUUG	4740
ACUUCAGCCU	AGACCCAACC	UUCACCAUCA	CCACUCAAAC	CGUCCCUCAG	GACGCUGUCU	4800
CCCGUAGUCA	ACGUAGAGGG	AGAACUGGGA	GGGGGCGAUU	GGGCRUUUAC	AGGUAUGUUU	4860
CGUCAGGYGA	RRGGCCGUCU	GGGAUGUUCG	ACAGCGUAGU	GCYCUGCGAG	UGCUAUGAUG	4920
CCGGGGCAGC	CUGGUACGAG	CUUACACCUG	CUGAGACUAC	GGUGAGACUC	CGGGCYUAUU	4980
UCAACACGCC	CGGUUUGCCC	GUAUGUCAAG	ACCACCUGGA	GUUCUGGGAA	GCGGUCUUUA	5040
CAGGUCUCAC	WCACAUURAC	GCCCACUUC	UCUCCCAGAC	GAAGCAAGGA	GGAGAAAACU	5100
UUGCRUAUCU	AACGGCCUAC	CAGGCCACAG	UAUGCGCCAG	GGCAAAGGCC	CCUCCUCCUU	5160
CGUGGGACGU	GAUGUGGAAG	UGUCUAACUA	GGCUCAAACC	UACACUGACU	GGUCCCACCC	5220
CCCUCUGUA	CCGCUUGGGU	GCCGUGACCA	AUGAGGUYAC	CUUGACGCAC	CCCGUGACGA	5280
AAUACAUCGC	CACGUGCAUG	CAAGCUGACC	UYGAGAUCAU	GACAAGCUC	UGGGUCCUGG	5340
CGGGGGGGGU	GCUAGCCGCC	GUGGCAGCUU	ACUGCCUGGC	GACUGGCUGC	AUUUCCAUCA	5400
UUGGCCGCCU	ACACCUGAAU	GAUCGGGUGG	UUGUGRCCCC	YGACAAGGAR	AUCUUAUAUG	5460
AGGCCUUUGA	UGAGAUGGAA	GAAUGCGCCU	CCAAAGCCGC	CCUCAUUGAG	GAAGGGCAGC	5520
GGAUGGCGGA	GAUGCUCAAA	UCUAAGAUAC	AAGGCCUCCU	ACAACAGGCC	ACAAGGCAAG	5580
CUCAAGRCAU	RCAGCCAGCU	AUACAGUCAU	CAUGGCCCAA	GCUUGAACAA	UUUUGGGCCA	5640
AACACAUGUG	GAACUUCAUC	AGUGGUUAUC	AGUACCUAGC	AGGACUCUCC	ACCCUACCGG	5700
GAAAUCCUGC	AGURGCAUCA	AUGAUGGCUU	UUAGCGCCGC	GCUGACUAGC	CCACUACCCA	5760
CCAGCACCAC	CAUCCUCUUG	AACAUCAUGG	GAGGAUGCUU	GGCCUCYCAG	AUUGCCCCCC	5820
CUGCCGGAGC	CACYGGCUUC	GUUGUCAGUG	GUCUAGUGGG	GGCGGCCGUC	GGAAGCAUAG	5880
GCCUGGGUAA	GAUACUGGUG	GACGUUUUGG	CCGGGUACGG	CGCAGGCAUU	UCAGGGGCCC	5940
UCGUAGCUUU	UAAGAUCAUG	AGCGGCGAGA	AGCCACGGU	AGAAGACGUU	GUGAAUCUCC	6000
UGCCUGCUAU	YCUGUCUCCU	GGUGCGYUGG	UAGUGGGAGU	CAUCUGUGCA	GCAAUYCUGC	6060
GCCGCCACGU	CGGUCAGGGA	GAGGGRGCGG	UCCAGUGGAU	GAACAGACUG	AUCGCCUUCG	6120
CCUCCAGGGG	AAACCACGUU	GCCCCUACCC	ACUACGUGGU	GGAGUCUGAC	GCUUACACGC	6180
GUGURACGCA	GGUGCUGAGU	UCACUUACAA	UUACCAGCUU	ACUUAGGAGA	CUACAUGCCU	6240
GGAUCACUGA	AGAUUGCCCA	RUCCCAUGCU	CGGGGUCUUG	GCUCCAGGAC	AUUUGGGAAU	6300
GGGUUUUGUUC	CAUCCUCACA	GACUUYAAAA	ACUGGCUGUC	UUCAAAAUUA	CUCCCCAAGA	6360

UGCCCCGGCAU UCCCUUUUUAUC UCUUGCCAGA AGGGAUACAA GGGUGUAUGG GCUGGUACGG 6420
 GUGUCAUGAC YACUCGRURC CCAUGUGGAG CAAACAUCUC GGGCCAUGUC CGCAUGGGCA 6480
 CCAUGAAAAU AACAGGCCCCG AAGACUUGCU UGAACCUUGUG GCAGGGGACU UCCCCCAUUA 6540
 AUUGUUACAC AGAAGGGCCY UGCGUGCCAA AACCCCUCC UAAUUACAAG ACCGCAUUU 6600
 GGAGGGUGGC AGCGUCGGAG UACGUUGAGG UCACACAGCA UGGCUCUUUC UCGUAUGUAA 6660
 CRGGGUUAAC CAGUGACAAC CUUAAGGUYC CUUGCCAGGU ACCAGCUCCA GAAUUUUUCU 6720
 CUUGGGUGGA CGGGGUGCAA AUCCACCGAU UCGCCCCCGU WCCAGGUCCC UUCUUUCGGG 6780
 AUGAGGUAAAC GUUCACCGUA GGCCUUAACU CCUUCGUGGU CGGCUCUCAG CUCCCUUGCG 6840
 AUCCUGAGCC GGACACCGAR GUACUGGCCU CYAUGUUGAC AGACCCGUCC CACAUCACCG 6900
 CKGAGGCGGC AGCCAGGCGA UUGGCAAGGG GAUCUCCCC YUCACAGGCU AGCUCCUCAG 6960
 CGAGCCAGCU CUCUGCCCCG UCCUUGAAGG CUACCUGUAC CACCCAUAAG ACAGCAUAUG 7020
 AUUGUGACAU GGUGGAUGCY AACCUIIUUA UGGGAGGMGA UGUGAYCCGG AUUGAGUCUG 7080
 ACUCUAAGGU GAUCGUUCUA GACUCCUUC AUUCCAUGAC UGAGGUAGAG GAUGAUCGUG 7140
 AGCCUUCUGU ACCAUCAGAG UACCUGAUCA AGAGGAGAAA GUUCCACCG GCGCUGCCUC 7200
 CUUGGGCCCCG UCCAGACUAC AAUCCUGUUU UGAUCGAGAC AUGGAAGAGG CCGGGCUAUG 7260
 AACCACCCAC UGUCCUAGGC UGUGCCCUCC CCCCCACACU UCAAACGCCA GUGCCUCCAC 7320
 CUCGGAGGCG CCGCGCYAAA RUCCUGACCC AGGACRAUGU GGAGGGGRUC CUCAGGGAGA 7380
 UGGCUGACAA AGURCUCAGC CCUCUCCAAG ACAACAUGA CUCCGGUCAC UCCACUGGAG 7440
 CGGAUACCGG AGGAGACAUC GUCCAGCAAC CCUCUGACGA GACUGCCGCU UCAGAAGCGG 7500
 GGUCACUGUC CUCCAUGCCU CCCCUUGAGG GAGAGCCGGG AGACCCYGAC CUGGAGUUUG 7560
 AACCAGUGGG AUCCGCUCC CCUUCUGAGG GGGAGUGUGA GGUCAUUGAU UCGGACUCUA 7620
 AGUCGUGGUC CACAGUCUCU GAUCAAGAGG AUUCUGUUAU CUGCUGCUCU AUGUCAUACU 7680
 CCUGGACGGG GGCCCUCAUA ACACCAUGUG GGCCGAAGA GGAGAAGUUA CCGAUCAACC 7740
 CUCUGAGUAA UUCGCUCAUG CGGUUCCAUA AYAAGGUGUA CUCCACAACC UCGAGGAGUG 7800
 CCUCUCUGAG GGCAAAGAAG GUGACUUUG ACAGGGUGCA GGUGCUGGAC GCACACUAUG 7860
 ACUCAGUCUU GCAGGACGUU AAGCGGGCCG CCUCUAAGGU URGUGCGAGG CUCCUCACAG 7920
 UAGAGGAAGC CUGCGCGCUG ACCCCGCCCC ACUCCGCCAA AUCGCGAUAC GGAUUUGGGG 7980
 CAAAAGAGGU GCGCAGCUUA UCCAGGAGGG CCGUUAACCA CAUCCGGUCC GUGUGGGAGG 8040
 ACCUCCUGGA AGACCAACRU ACCCCAAUG ACACAACUUA CAUGGCUAAA AAUGAGGUGU 8100

UCUGCAUUGA UCCAACUAAR GGUGGGAAAA AGCCAGCUCG CCUCAUCGUA UACCCCGACC 8160
 UUGGGGUCAG GGUGUGCGAA AAGAUGGCCC UCUAUGACAU CRCACAAAAG CUUCCCAAAG 8220
 CGAUAAUGGG GCCAUCCUUAU GGGUUCCAAU ACUCUCCCGC AGAACGGGUC GAUUUCCUCC 8280
 UCAAAGCUUG GGGAAGUAAG AAGGACCCAA UGGGGUUCUC GUAUGACACC CGCUGCUUUG 8340
 ACUCAACCGU CACGGAGAGG GACAUAAAGAA CAGAAGAAUC CAUAUAUCAG GCUUGUUCUC 8400
 UGCCUCAAGA AGCCAGAACU GUCAUACACU CGCUCACUGA GAGACUUUAC GUAGGAGGGC 8460
 CCAUGACAAA CAGCAAAGGG CAAUCCUGCG GCUACAGGCG UUGCCGCGCA AGCGGKGUUU 8520
 UCACCACCAG CAUGGGGAAU ACCAUGACAU GUUACAUCAA AGCCCUUGCA GCGUGUAAGG 8580
 CUGCRGGGAU CGUGGACCCU GUUAUGUUGG UGUGUGGAGA CGACCUGGUC GUCAUCUCAG 8640
 AGAGCCAAGG UAACGAGGAG GACGAGCGAA ACCUGAGAGC UUUCACGGAG GCUAUGACCA 8700
 GGUAUUCCGC CCCUCCCGGU GACCUUCCCA GACCGGAUA UGACUUGGAG CUUAUAACAU 8760
 CCUGCUCCUC AAACGUUUCG GUAGCGCUGG ACUCUCGGGG UCGCCGCCGG UACUUCCUAA 8820
 CCAGAGACCC UACCACUCCA AUCACCCGAG CUGCUUGGGA AACAGUAAGA CACUCCCCUG 8880
 UCAAUUCUUG GCUGGGCAAC AUCAUCCAGU ACGCCCCAC AAUCUGGGUC CGGAUGGUCA 8940
 UAAUGACUCA CUUCUUCUCC AUACUAUUGG CCCAGGACAC UCUGAACCAA AAUCUCAAUU 9000
 UUGAGAUGUA CGGGGCAGUA UACUCGGUCA AUCCAUAAGA CCUACCGGCC AUAAUUGAAA 9060
 GGCUACAUGG GCUUGAAGCC UUUUCACUGC ACACAUACUC UCCCCACGAA CUCUCACGGG 9120
 UGGCAGCAAC UCUCAGAAAA CUUGGAGCGC CUCCCCUAG AGCGUGGAAG AGUCGGGCGC 9180
 GUGCCGUGAG AGCUUCACUC AUCGCCAAG GAGCGAGGGC GGCCAUUUGU GGCCGCUACC 9240
 UCUIICAACUG GGCGGUGAAA ACAAAGCUA AACUCACUCC AUUGCCCGAG GCGAGCCGCC 9300
 UGGAUUUAUC CGGGUGGUUC ACCGUGGGCG CCGCGGGGGG CGACAUUUAU CACAGCGUGU 9360
 CGCAUGCYCG ACCCCGCCUA UUACUCCUUU GCCUACUCCU ACUUAAGCGUA GGAGUAGGCA 9420
 UCUUUUUACU CCCCUCUCGG UAGAGCGGCA AACYCUAGCU ACACUCCAUA GCUAGUUUCC 9480
 GUUUUUUUUU UUUUUUUUUU UUUUUUUUUU U 9511

Sequence ID No. 7
 Sequence Length: 9.511
 Sequence Type: nucleic acid
 Strandedness: single
 Topology: linear
 Molecule Type: cDNA to genomic RNA
 Method for Determination of Feature: E

```

GCCCCCCCC TGATGGGGG GACACTCCG CATGAATCAC TCCCCTGTGA GGAACACTG 60
TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG TGTCGTACAG CCTCCAGGCC 120
CCCCCCTCCC GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTACCGG 180
AAAGACTGGG TCCTTTCTTG GATAAACCCA CTCTATGTCC GGTCAATTGG GCACGCCCCC 240
GCAAGACTGC TAGCCGAGTA GCGTTGGGTT GCGAAAGGCC TTGTGGTACT GCCTGATAGG 300
GTRCTTGCGA GTGCCCCGGG AGGTCTCGTA GACCGTGCAT CATGAGCACA AATCCTAAAC 360
CTCAAAGAAA AACCAAAAGA AACACAAACC GCGGCCACCA GGACGTTAAG TTCCCGGGTG 420
GCGGTCAGAT CGTTGGCGGA GTTTACTTGC TGCCGCGCAG GGGCCCCAGG TTGGGTGTGC 480
GCGCGACAAG GAAGACTTCY GAGCGATCCC AGCCGCGTGG ACGACGCCAG CCCATCCCGA 540
AAGATCGGCG CTCCACCGGC AAGTCCTGGG GAAAGCCAGG ATATCCTTGG CCCCTGTACG 600
GAAACGAGGG TTGCGGCTGG GCGGGTTGGC TCCTGTCCCC CCGCGGGTCT CGTCCTACTT 660
GGGGCCCCAC CGACCCCCGG CATAGATCAC GCAATTTGGG CAGAGTCATC GATACCATT 720
CGTGTGGTTT TGCCGACCTC ATGGGGTACA TCCCTGTCTG TGGCGCCCCG GTYGGAGGCG 780
TCGCCAGAGC TCTGGCACAC GGTGTTAGGG TCCTGGAGGA CGGGATAAAT TACGCAACAG 840
GGAATTTACC CGGTTGCTCT TTTTCTATCT TTTTGTCTGC TCTTCTGTCA TCGGTCACAR 900
TGCCAGTGTC TGCAGTGGAA GTCAGGAACA TYAGTTCTAG CTACTACGCC ACTAATGATT 960
GCTCAAACAA CAGCATCACC TGGCAGCTCA CTGACGCAGT TCTCCATCTT CCTGGATGCG 1020
TCCCATGTGA GAAYGATAAY GGCACCTTGC RTTGCTGGAT ACAAGTAACA CCCRACGTGG 1080
CTGTGAAACA CCGCGGTGCG CTCACTCGTA GCCTGCGAAC ACACGTCGAC ATGATCGTAA 1140
TGGCAGCTAC GGCCTGCTCG GCCTTGTATG TGGGAGATGT GTGCGGGGCC GTGATGATYC 1200

```

TATCGCAGGC TTTCATGGTA TCACCACAAC GCCACAACCT CACCCAAGAG TGCAACTGTT 1260
 CCATCTACCA AGGTCACATC ACCGGGCCATC GCATGGCATG GGACATGATG CTRARCTGGT 1320
 CTCCAACTCT TRCCATGATC CTCGCCTACG CYGCTCGYGT TCCCGARCTG GTCCTCGAAA 1380
 TYATYTTGGG CGGCCATTGG GGTGTGGYGT TYGGCTTGGG CTATTTCTCC ATGCARGGAG 1440
 CGTGGGCCAA AGTCRTYGCC ATCCTCCTTC TTGTTGCGGG AGTGGATGCA WCCACCTATT 1500
 CCASCGGYCA GSAAGCGGGT CGTRCCGYCK HKGGGWTKC TRGCCTCTTT AMTACTGGTG 1560
 CCAAGCAGAA CCTCYATTTR ATCAACACCA ATGGCAGCTG GCACATAAAC CGGACTGCCC 1620
 TCAATTGCAA TGACAGCYTA SAGACGGGTT TCHTCGCTTC CYTGKTTTAC WMCCRCARGT 1680
 TCAACAGCTC TGGCTGCCCC GAGCGCTTGT CTCCTGCGG CGGGCTGGAC GAYTTYCGCA 1740
 TCGGCTGGGG AACCTTGGA TACGAAACCA ACGTACCAA CGATGRGGAC ATGAGGCCGT 1800
 ACTGCTGGCA TTACCCCCCG AGGCCCTTGG GCATCGTCCC GGCTAGGACG GTTTGCGGAC 1860
 CGGTCTATTG YTTACCCCT AGCCCTGTTG TCGTGGGCAC CACTGACAAG CAGGGCGTAC 1920
 CCACCTACAC CTGGGGRGAA AACGAGACCG ATGTCTTCT GCTRAATAGC ACAAGACCCC 1980
 CGCGAGGAGC TTGGTTCGGC TGCACYTGA TGAACGGGAC TGGGTTCAC AAGACATGCG 2040
 GTGCACCACC TTGCCGCATT AGGAAAGACT ACAACAGCAC TCTCGATTTA TTGTGCCCCA 2100
 CAGACTGTTT TAGGAAGCAC CCAGATGCTA CCTATCTTAA GTGTGGAGCA GGGCCTTGGT 2160
 TAACTCCAG GTGCCTGGTA GACTACCCTT ATAGRYTGTG GCATTATCCG TGCCTGTAA 2220
 ACTTCACCAT CTTAAGGCG CGGATGTATG TAGGAGGGGT GGAGCATCGA TTCTCCGCAG 2280
 CATGCAACTT CACGCGCGGA GATCGCTGCA GACTGGAAGA TAGGGATAGG GGYCAGCAGA 2340
 GTCCACTGCT GCATTCCACT ACTGAGTGGG CGGTGYTCCC ATGCTCCTTC TCTGACCTAC 2400
 CAGCACTATC CACTGGCCTA TTGCACCTCC ACCAAAACAT CGTGGACGTG CAGTACCTYT 2460
 ACGGACTTTC TCCGGCTCTG ACAAGATACA TCGTGAAGTG GGAGTGGGTG ATCCTCCTTT 2520
 TCTTGTGTT GGCAGACGCC AGGRTCTGTG CATGCCTTTG GATGCTCAWC ATACTGGGCC 2580
 AAGCCGAAGC GGCCTTGAG AAGCTCATCA TCTTGCACTC CGCTAGYGCT GCTAGTGCCA 2640
 ATGGTCCGCT GTGGTTTTTC ATCTTCTTTA CAGCGGCCTG GACTTTAAAG GGCAGGGTGG 2700
 TCCCCGTGGC CACGTA CTCT GTBCTCGGCT TRTGGTCCTT CCTCCTCCTA GTCCTGGCYT 2760
 TACCACAGCA GGCTTATGCC TTGGACGCTG CTGAACAAGG GGAAGTGGGG CTGGCCATAT 2820
 TAGTAATTAT ATCCATCTTT ACTCTTACCC CAGCATACAA GATCCTCCTG AGCCGTTTAC 2880
 TGTGGTGGCT GTCCTACATG CTGGTCTTGG CCGAGGCCCA GATTACAGCA TGGGTTCCCC 2940

CCCTGGAGGT CCGAGGGGGG CGTGACGGGA TCATCTGGGT GGCTGTCAAT CTACACCCAC 3000
 GCCTTGTTGT TGAGGTCACG AAATGGTTGT TAGCAATCCT GGGGCCTGCC TACCTCCTTA 3060
 RAGCGTCTCT GCTACGGATA CCGTACTTTG TGAGGGCCCA CGCTTTGCTA CGAGTGTGTA 3120
 CCCTGGTGAA ACACCTCGCR GGGGCTAGGT ACATCCAGAT GCTGTTRATC ACCATAGGCA 3180
 GATGGACCGG CACTTACATC TACGACCACC TCTCCCTTTT ATCAACTTGG GCGGCCAGG 3240
 GTTTRCGGGA CCTGGCAATC GCCGTGGAGC CTGTGGTGTT CAGCCCAATG GAGAAGAAGG 3300
 TCATTGTGTG GGGGGCTGAG ACAGTGGCGT GTGGAGACAT CCTGCATGGC CTCCCGGTCT 3360
 CCGCGAGGCT AGGTAGGGAR GTTCTGCTCG GCCCTGCCGA CGGCTACACC TCCAAGGGGT 3420
 GGAAKCTCCT AGCTCCCAT TCTGCTTACA CTCAGCAAAC TCGTGGTCTC CTGGGTGCTA 3480
 TCGTGGTCAG CCTAACGGGC CGCGACAAAA ATGAGCAGGC TGGGCAGGTC CAGGTTCTGT 3540
 CCTCCGTAC ACAAACTTTC TTGGGGACAT CCATTTGCGG CGTCCTCTGG ACAGTATATC 3600
 ACGGGGCTGG TAATAAGACC TTGGCCGGCC CCAAGGGACC AGTCACTCAG ATGTACACCA 3660
 GCGCAGAAAG GGACCTCGTG GGATGGCCTA GTCCCCCGG GACTAAGTCA TTGGACCCCT 3720
 GTACCTGGGG GGCCGTAGAC CTCTACCTGG TCACCCGAAA CGCTGATGTC ATTCCGGTCC 3780
 GGAGGAAAGA TGACCGACGG GGTGCATTAC TCTCGCCAAG GCCCTCTCA ACCCTCAAAG 3840
 GATCATCCGG AGGGCCCGTG CTCTGCTCWA GGGGACACGC CGTGGGCTTG TTCAGAGCGG 3900
 CCGTGTGTGC CAGGGGTGTA GCCAAATCTA TTGACTTCAT CCCCCTCGAA TCACTCGATR 3960
 TCGCCACACG GACGCCCAGT TTCTCTGACA ACAGTRCGCC GCCAGCTGTG CCCAGTCTT 4020
 ACCAGTGGG TTAATTGCAC GCACCAACAG GCAGCGGAAA GAGCACCAG GTCCCTGCCG 4080
 CGTATGCCAG TCAGGGGTAT AAAGTACTCG TACTAAATCC CTCTGTGCGG GCCACACTTG 4140
 GTTTTGGGGC CTACATGTCC AAAGCCCACG GGATCAACCC TAATATCAGA ACTGGAGTGC 4200
 GGACCGTTAC CACCGGGGAC TCTATCACTT ACTCCACTTA TGGCAAGTTT ATCGCAGATG 4260
 GAGGCTGTGC AGCCGGTGCC TATGACATCA TCATATGCGA CGAATGCCAT TCAGTGGACG 4320
 CTACTACCAT CCTTGGCATT GGAACAGTCC TTGACCAAGC TGAGACCGCA GCGTCAGGC 4380
 TAGTGGTYTT GGCCACAGCC ACGCTCCCG GTACGGTGAC AACTCCCCAC AGTAACATAG 4440
 AGGAGGTGGC CCTTGGTCAC GAGGGCGAGA TCCCTTTTTA TGGCAAAGCT ATTCCCCTAG 4500
 CTTTCATCAA GGGGGGCAGA CACTTGATCT TTTGCCATTC AAAGAAGAAG TGCGACGAGC 4560
 TCGCAGCGGC CCTCCGGGGC AYGGGTGTCA ATGCCGTTGC ATACTATAGG GGTCTCGACG 4620
 TCTCCGTTAT ACCAACTCAA GGAGACGTGG TGGTTGTGCG CACTGATGCC CTAATGACTG 4680

GGTACACCGG CGACTTTGAC TCGTCATCG ACTGTAATGT TGCAGTCTCT CAGATTGTTG 4740
ACTTCAGCCT AGACCCAACC TTCACCATCA CCACTCAAAC CGTCCCTCAG GACGCTGTCT 4800
CCCGTAGTCA ACGTAGAGGG AGAACTGGGA GGGGGCGATT GGGCRTTTAC AGGTATGTTT 4860
CGTCAGGYGA RRGGCCGTCT GGGATGTTG ACAGCGTAGT GCYCTGCGAG TGCTATGATG 4920
CCGGGGCAGC CTGGTACGAG CTTACACCTG CTGAGACTAC GGTGAGACTC CGGGCYTATT 4980
TCAACACGCC CGGTTTGCCC GTATGTCAAG ACCACCTGGA GTTCTGGGAA GCGGTCTTTA 5040
CAGGTCTCAC WCACATTRAC GCCCACTTCC TCTCCCAGAC GAAGCAAGGA GGAGAAACT 5100
TTGCRATCT AACGGCCTAC CAGGCCACAG TATGCGCCAG GGCAAAGGCC CCTCCTCCTT 5160
CGTGGGACGT GATGTGGAAG TGTCTAACTA GGCTCAAACC TACACTGACT GGTCCCACCC 5220
CCCTCCTGTA CCGCTTGGGT GCCGTGACCA ATGAGGTIAC CTTGACGCAC CCCGTGACGA 5280
AATACATCGC CACGTGCAIG CAAGCTGACC TYGAGATCAT GACAAGCTCA TGGGTCTTGG 5340
CGGGGGGGGT GCTAGCCGCC GTGGCAGCTT ACTGCCCTGGC GACTGGCTGC ATTTCCATCA 5400
TTGGCCGCCT ACACCTGAAT GATCGGGTGG TTGTGRCCCC YGACAAGGAR ATCTTATATG 5460
AGGCCCTTGA TGAGATGGAA GAATGCGCCT CCAAAGCCGC CCTCATTGAG GAAGGGCAGC 5520
GGATGGCGGA GATGCTCAAA TCTAAGATAC AAGGCCTCCT ACAACAGGCC ACAAGGCAAG 5580
CTCAAGRCAT RCAGCCAGCT ATACAGTCAT CATGGCCCAA GCTTGAACAA TTTTGGGCCA 5640
AACACATGTG GAACTTCATC AGTGGTATAC AGTACCTAGC AGGACTCTCC ACCCTACCGG 5700
GAAATCCTGC AGTRGCATCA ATGATGGCTT TTAGCGCCGC GCTGACTAGC CCACTACCCA 5760
CCAGCACCAAC CATCCTCTTG AACATCATGG GAGGATGCTT GGCCTCYCAG ATTGCCCCCC 5820
CTGCCGGAGC CACYGGCTTC GTTGTCAGTG GTCTAGTGGG GCGGGCCGTC GGAAGCATAG 5880
GCCTGGGTAA GATACTGGTG GACGTTTTGG CCGGGTACGG CGCAGGCATT TCAGGGGCCC 5940
TCGTAGCTTT TAAGATCATG AGCGGCGAGA AGCCACCGGT AGAAGACGTT GTGAATCTCC 6000
TGCTGTCTAT YCTGTCTCCT GGTGCGYTGG TAGTGGGAGT CATCTGTGCA GCAATYCTGC 6060
GCCGCCACGT CGGTCAGGGA GAGGGRGCGG TCCAGTGGAT GAACAGACTG ATCGCCTTCG 6120
CCTCCAGGGG AAACCACGTT GCCCCTACCC ACTACGTGGT GGAGTCTGAC GCTTCACAGC 6180
GTGTRACGA GGTGCTGAGT TCACTTACAA TTACCAGCTT ACTTAGGAGA CTACATGCCT 6240
GGATCACTGA AGATTGCCCA RTCCCATGCT CGGGGTCTTG GCTCCAGGAC ATTTGGGATT 6300
GGGTTTGTTT CATCCTCACA GACTTYAAAA ACTGGCTGTC TTCAAATA CTCCCCAAGA 6360
TGCCCGGCAT TCCCTTTATC TCTTGCCAGA AGGGATACAA GGGTGTATGG GCTGGTACGG 6420

GTGTCATGAC YACTCGRTRC CCATGTGGAG CAAACATCTC GGGCCATGTC CGCATGGGCA 340
 CCATGAAAAT AACAGGCCCG AAGACTTGCT TGAACCTGTG GCAGGGGACT TTCCCCATTA 354
 ATTGTTACAC AGAAGGGCCY TCGTGCCAA AACCCCTCC TAATTACAAG ACCGCAATTT 360
 GGAGGGTGGC AGCGTCGGAG TACGTTGAGG TCACACAGCA TGGCTCTTC TCGTATGTAA 366
 CRGGGTAAAC CAGTGACAAC CTTAAGGTTC CTTGCCAGGT ACCAGCTCCA GAATTTTTCT 372
 CTTGGGTGGA CGGGGTGCAA ATCCACCGAT TCGCCCCGT WCCAGGTCCC TTCITTCGGG 378
 ATGAGGTAAC GTTCACCGTA GGCCTTAACT CCTTCGTGGT CGGCTCTCAG CTCCCTTGCG 384
 ATCCTGAGCC GGACACCGAR GTACTGGCCT CYATGTTGAC AGACCCGTCC CACATCACCG 390
 CKGAGGGGGC AGCCAGGCCA TTGGCAAGGG GATCTCCCC YTCACAGGT AGCTCCTCAG 396
 CGAGCCAGCT CTCTGCCCCG TCCTTGAAGG CTACCTGTAC CACCCATAAG ACAGCATATG 702
 ATTGTGACAT GGTGGATGCY AACCTTTTCA TGGGAGGMGA TGTGAYCCGG ATTGAGTCTG 708
 ACTCTAAGGT GATCGTTCTA GACTCCCTCG ATTCCATGAC TGAGGTAGAG GATGATCGTG 714
 AGCCTTCTGT ACCATCAGAG TACCTGATCA AGAGGAGAAA GTTCCCACCG GCGCTGCCTC 720
 CTTGGGCCCCG TCCAGACTAC AATCCTGTTT TGATCGAGAC ATGGAAGAGG CCGGGCTATG 726
 AACCACCCAC TGTCTAGGC TGTGCCCTCC CCCCCACAC TCAAACGCCA GTGCCTCCAC 732
 CTCGGAGGCG CCGCGCYAAA RTCCTGACCC AGGACRATGT GGAGGGGRTC CTCAGGGAGA 738
 TGGCTGACAA AGTRCTCAGC CCTCTCCAAG ACAACAATGA CTCCGGTCAC TCCACTGGAG 744
 CGGATACCGG AGGAGACATC GTCCAGCAAC CCTCTGACGA GACTGCCGCT TCAGAAGCGG 750
 GGTCACTGTC CTCCATGCCT CCCCTTGAGG GAGAGCCGGG AGACCCYGAC CTGGAGTTTG 756
 AACCAGTGGG ATCCGCTCCC CCTTCTGAGG GGGAGTGTGA GGTCAATTGAT TCGGACTCTA 762
 AGTCGTGGTC CACAGTCTCT GATCAAGAGG ATTCTGTTAT CTGCTGCTCT ATGTCATACT 768
 CCTGGACGGG GGCCCTCATA ACACCATGTG GGCCCGAAGA GGAGAAGTTA CCGATCAACC 774
 CTCTGAGTAA TTCGCTCATG CGGTTCCATA AYAAGGTGTA CTCCACAACC TCGAGGAGTG 780
 CCTCTCTGAG GGCAAAGAAG GTGACTTTTG ACAGGGTGCA GGTGCTGGAC GCACACTATG 786
 ACTCAGTCTT GCAGGACGTT AAGCGGGCCG CCTCTAAGGT TRGTGCGAGG CTCCTCACAG 792
 TAGAGGAAGC CTGCGCGCTG ACCCGCCCC ACTCCGCCAA ATCGCGATAC GGATTTGGGG 798
 CAAAAGAGGT GCGCAGCTTA TCCAGGAGGG CCGTTAACCA CATCCGGTCC GTGTGGGAGG 804
 ACCTCCTGGA AGACCAACRT ACCCAATTG ACACAACAT CATGGCTAAA AATGAGGTGT 810
 TCTGCATTGA TCCAACAAAR GGTGGGAAAA AGCCAGCTCG CCTCATCGTA TACCCCGACC 816

TTGGGGTCAG GGTGTGCGAA AAGATGGCCC TCTATGACAT CRCACAAAAG CTTCCTCAAAG 8220
 CGATAATGGG GCCATCCTAT GGGTTCCAAT ACTCTCCCGC AGAACGGGTC GATTTCTCTC 8280
 TCAAAGCTTG GGGAAGTAAG AAGGACCCAA TGGGGTTCTC GTATGACACC CGCTGCTTTG 8340
 ACTCAACCGT CACGGAGAGG GACATAAGAA CAGAAGAATC CATATATCAG GCTTGTTC 8400
 TGCCTCAAGA AGCCAGAACT GTCATACACT CGCTCACTGA GAGACTTTAC GTAGGAGGGC 8460
 CCATGACAAA CAGCAAAGGG CAATCCTGCG GCTACAGGCG TTGCCGCGCA AGCGGKGT 8520
 TCACCACCAG CATGGGGAAT ACCATGACAT GTTACATCAA AGCCCTTGCA GCGTGTAAGG 8580
 CTGCRGGGAT CGTGGACCCT GTTATGTTGG TGTGTGGAGA CGACCTGGTC GTCATCTCAG 8640
 AGAGCCAAGG TAACGAGGAG GACGAGCGAA ACCTGAGAGC TTTCACGGAG GCTATGACCA 8700
 GGTATTCCGC CCTCCCGGT GACCTTCCCA GACCGGAATA TGAATTGGAG CTTATAACAT 8760
 CCTGCTCCTC AAACGTATCG GTAGCGCTGG ACTCTCGGGG TCGCCGCCGG TACTTCCIAA 8820
 CCAGAGACCC TACCACTCCA ATCACCCGAG CTGCTTGGGA AACAGTAAGA CACTCCCCTG 8880
 TCAATTCTTG GCTGGGCAAC ATCATCCAGT ACGCCCCAC AATCTGGGTC CGGATGGTCA 8940
 TAATGACTCA CTCTTCTCC ATACTATTGG CCCAGGACAC TCTGAACCAA AATCTCAATT 9000
 TTGAGATGTA CGGGGCAGTA TACTCGGTCA ATCCATTAGA CCTACCGGCC ATAATTGAAA 9060
 GGCTACATGG GCTTGAAGCC TTTTCACTGC ACACATACTC TCCCCACGAA CTCTCACGGG 9120
 TGGCAGCAAC TCTCAGAAAA CTGGAGCGC CTCCCCTTAG AGCGTGGAAG AGTCGGGCGC 9180
 GTGCCGTGAG AGCTTCACTC ATCGCCCAAG GAGCGAGGGC GGCCATTTGT GGCCGCTACC 9240
 TCTTCAACTG GGCGGTGAAA ACAAAGCTCA AACTCACTCC ATTGCCCGAG GCGAGCCGCC 9300
 TGGATTTATC CGGGTGGTTC ACCGTGGGCG CCGGCGGGGG CGACATTTAT CACAGCGTGT 9360
 CGCATGCYCG ACCCGCCTA TTA CTCTCTTT GCCTACTCCT ACTTAGCGTA GGAGTAGGCA 9420
 TCTTTTTACT CCGCGCTCGG TAGAGCGGCA AACYCTAGCT AACTCCATA GCTAGTTTCC 9480
 GTTTTTTTTT TTTTTTTTTT TTTTTTTTTT T 9511

Sequence ID No. 8
 Sequence Length: 3,033
 Sequence Type: amino acid
 Topology: linear
 Molecule Type: protein

Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn	Thr
				5					10					15
Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile
				20					25					30
Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	Leu	Gly
				35					40					45
Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	Pro	Arg	Gly
				50					55					60
Arg	Arg	Gln	Pro	Ile	Pro	Lys	Asp	Arg	Arg	Ser	Thr	Gly	Lys	Ser
				65					70					75
Trp	Gly	Lys	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly
				80					85					90
Cys	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro
				95					100					105
Thr	Trp	Gly	Pro	Thr	Asp	Pro	Arg	His	Arg	Ser	Arg	Asn	Leu	Gly
				110					115					120
Arg	Val	Ile	Asp	Thr	Ile	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly
				125					130					135
Tyr	Ile	Pro	Val	Val	Gly	Ala	Pro	Val	Gly	Gly	Val	Ala	Arg	Ala
				140					145					150
Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Ile	Asn	Tyr	Ala
				155					160					165

Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala		
170	175	180
Leu Leu Ser Cys Val Thr Val Pro Val Ser Ala Val Glu Val Arg		
185	190	195
Asn Ile Ser Ser Ser Tyr Tyr Ala Thr Asn Asp Cys Ser Asn Asn		
200	205	210
Ser Ile Thr Trp Gln Leu Thr Asp Ala Val Leu His Leu Pro Gly		
215	220	225
Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu His Cys Trp Ile		
230	235	240
Gln Val Thr Pro Asn Val Ala Val Lys His Arg Gly Ala Leu Thr		
245	250	255
Arg Ser Leu Arg Thr His Val Asp Met Ile Val Met Ala Ala Thr		
260	265	270
Ala Cys Ser Ala Leu Tyr Val Gly Asp Val Cys Gly Ala Val Met		
275	280	285
Ile Leu Ser Gln Ala Phe Met Val Ser Pro Gln Arg His Asn Phe		
290	295	300
Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly His Ile Thr Gly		
305	310	315
His Arg Met Ala Trp Asp Met Met Leu Ser Trp Ser Pro Thr Leu		
320	325	330
Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro Glu Leu Val Leu		
335	340	345
Glu Ile Ile Phe Gly Gly His Trp Gly Val Val Phe Gly Leu Ala		
350	355	360
Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val Ile Ala Ile Leu		
365	370	375
Leu Leu Val Ala Gly Val Asp Ala Thr Thr Tyr Ser Ser Gly Gln		

380	385	390
Glu Ala Gly Arg Thr Val Ala Gly Phe	Ala Gly Leu Phe Thr Thr	
395	400	405
Gly Ala Lys Gln Asn Leu Tyr Leu Ile	Asn Thr Asn Gly Ser Trp	
410	415	420
His Ile Asn Arg Thr Ala Leu Asn Cys	Asn Asp Ser Leu Gln Thr	
425	430	435
Gly Phe Leu Ala Ser Leu Phe Tyr Thr	His Lys Phe Asn Ser Ser	
440	445	450
Gly Cys Pro Glu Arg Leu Ser Ser Cys	Arg Gly Leu Asp Asp Phe	
455	460	465
Arg Ile Gly Trp Gly Thr Leu Glu Tyr	Glu Thr Asn Val Thr Asn	
470	475	480
Asp Gly Asp Met Arg Pro Tyr Cys Trp	His Tyr Pro Pro Arg Pro	
485	490	495
Cys Gly Ile Val Pro Ala Arg Thr Val	Cys Gly Pro Val Tyr Cys	
500	505	510
Phe Thr Pro Ser Pro Val Val Val Gly	Thr Thr Asp Lys Gln Gly	
515	520	525
Val Pro Thr Tyr Thr Trp Gly Glu Asn	Glu Thr Asp Val Phe Leu	
530	535	540
Leu Asn Ser Thr Arg Pro Pro Arg Gly	Ala Trp Phe Gly Cys Thr	
545	550	555
Trp Met Asn Gly Thr Gly Phe Thr Lys	Thr Cys Gly Ala Pro Pro	
560	565	570
Cys Arg Ile Arg Lys Asp Tyr Asn Ser	Thr Ile Asp Leu Leu Cys	
575	580	585
Pro Thr Asp Cys Phe Arg Lys His Pro	Asp Ala Thr Tyr Leu Lys	
590	595	600

Cys Gly Ala Gly Pro Trp Leu Thr Pro Arg Cys Leu Val Asp Tyr		
605	610	615
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile		
620	625	630
Phe Lys Ala Arg Met Tyr Val Gly Gly Val Glu His Arg Phe Ser		
635	640	645
Ala Ala Cys Asn Phe Thr Arg Gly Asp Arg Cys Arg Leu Glu Asp		
650	655	660
Arg Asp Arg Gly Gln Gln Ser Pro Leu Leu His Ser Thr Thr Glu		
665	670	675
Trp Ala Val Leu Pro Cys Ser Phe Ser Asp Leu Pro Ala Leu Ser		
680	685	690
Thr Gly Leu Leu His Leu His Gln Asn Ile Val Asp Val Gln Tyr		
695	700	705
Leu Tyr Gly Leu Ser Pro Ala Leu Thr Arg Tyr Ile Val Lys Trp		
710	715	720
Glu Trp Val Ile Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Ile		
725	730	735
Cys Ala Cys Leu Trp Met Leu Ile Ile Leu Gly Gln Ala Glu Ala		
740	745	750
Ala Leu Glu Lys Leu Ile Ile Leu His Ser Ala Ser Ala Ala Ser		
755	760	765
Ala Asn Gly Pro Leu Trp Phe Phe Ile Phe Phe Thr Ala Ala Trp		
770	775	780
Tyr Leu Lys Gly Arg Val Val Pro Val Ala Thr Tyr Ser Val Leu		
785	790	795
Gly Leu Trp Ser Phe Leu Leu Leu Val Leu Ala Leu Pro Gln Gln		
800	805	810
Ala Tyr Ala Leu Asp Ala Ala Glu Gln Gly Glu Leu Gly Leu Ala		

815	820	825
Ile Leu Val Ile Ile Ser Ile Phe Thr Leu Thr Pro Ala Tyr Lys		
830	835	840
Ile Leu Leu Ser Arg Ser Val Trp Trp Leu Ser Tyr Met Leu Val		
845	850	855
Leu Ala Glu Ala Gln Ile Gln Gln Trp Val Pro Pro Leu Glu Val		
860	865	870
Arg Gly Gly Arg Asp Gly Ile Ile Trp Val Ala Val Ile Leu His		
875	880	885
Pro Arg Leu Val Phe Glu Val Thr Lys Trp Leu Leu Ala Ile Leu		
890	895	900
Gly Pro Ala Tyr Leu Leu Lys Ala Ser Leu Leu Arg Ile Pro Tyr		
905	910	915
Phe Val Arg Ala His Ala Leu Leu Arg Val Cys Thr Leu Val Lys		
920	925	930
His Leu Ala Gly Ala Arg Tyr Ile Gln Met Leu Leu Ile Thr Ile		
935	940	945
Gly Arg Trp Thr Gly Thr Tyr Ile Tyr Asp His Leu Ser Pro Leu		
950	955	960
Ser Thr Trp Ala Ala Gln Gly Leu Arg Asp Leu Ala Ile Ala Val		
965	970	975
Glu Pro Val Val Phe Ser Pro Met Glu Lys Lys Val Ile Val Trp		
980	985	990
Gly Ala Glu Thr Val Ala Cys Gly Asp Ile Leu His Gly Leu Pro		
995	1000	1005
Val Ser Ala Arg Leu Gly Arg Glu Val Leu Leu Gly Pro Ala Asp		
1010	1015	1020
Gly Tyr Thr Ser Lys Gly Trp Lys Leu Leu Ala Pro Ile Thr Ala		
1025	1030	1035

Tyr Thr Gln Gln Thr Arg Gly Leu Leu Gly Ala Ile Val Val Ser	1040	1045	1050
Leu Thr Gly Arg Asp Lys Asn Glu Gln Ala Gly Gln Val Gln Val	1055	1060	1065
Leu Ser Ser Val Thr Gln Thr Phe Leu Gly Thr Ser Ile Ser Gly	1070	1075	1080
Val Leu Trp Thr Val Tyr His Gly Ala Gly Asn Lys Thr Leu Ala	1085	1090	1095
Gly Pro Lys Gly Pro Val Thr Gln Met Tyr Thr Ser Ala Glu Gly	1100	1105	1110
Asp Leu Val Gly Trp Pro Ser Pro Pro Gly Thr Lys Ser Leu Asp	1115	1120	1125
Pro Cys Thr Cys Gly Ala Val Asp Leu Tyr Leu Val Thr Arg Asn	1130	1135	1140
Ala Asp Val Ile Pro Val Arg Arg Lys Asp Asp Arg Arg Gly Ala	1145	1150	1155
Leu Leu Ser Pro Arg Pro Leu Ser Thr Leu Lys Gly Ser Ser Gly	1160	1165	1170
Gly Pro Val Leu Cys Ser Arg Gly His Ala Val Gly Leu Phe Arg	1175	1180	1185
Ala Ala Val Cys Ala Arg Gly Val Ala Lys Ser Ile Asp Phe Ile	1190	1195	1200
Pro Val Glu Ser Leu Asp Val Ala Thr Arg Thr Pro Ser Phe Ser	1205	1210	1215
Asp Asn Ser Thr Pro Pro Ala Val Pro Gln Ser Tyr Gln Val Gly	1220	1225	1230
Tyr Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro	1235	1240	1245
Ala Ala Tyr Ala Ser Gln Gly Tyr Lys Val Leu Val Leu Asn Pro			

1250	1255	1260
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala		
1265	1270	1275
His Gly Ile Asn Pro Asn Ile Arg Thr Gly Val Arg Thr Val Thr		
1280	1285	1290
Thr Gly Asp Ser Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Ile Ala		
1295	1300	1305
Asp Gly Gly Cys Ala Ala Gly Ala Tyr Asp Ile Ile Ile Cys Asp		
1310	1315	1320
Glu Cys His Ser Val Asp Ala Thr Thr Ile Leu Gly Ile Gly Thr		
1325	1330	1335
Val Leu Asp Gln Ala Glu Thr Ala Gly Val Arg Leu Val Val Leu		
1340	1345	1350
Ala Thr Ala Thr Pro Pro Gly Thr Val Thr Thr Pro His Ser Asn		
1355	1360	1365
Ile Glu Glu Val Ala Leu Gly His Glu Gly Glu Ile Pro Phe Tyr		
1370	1375	1380
Gly Lys Ala Ile Pro Leu Ala Phe Ile Lys Gly Gly Arg His Leu		
1385	1390	1395
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Ala		
1400	1405	1410
Leu Arg Gly Met Gly Val Asn Ala Val Ala Tyr Tyr Arg Gly Leu		
1415	1420	1425
Asp Val Ser Val Ile Pro Thr Gln Gly Asp Val Val Val Val Ala		
1430	1435	1440
Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val		
1445	1450	1455
Ile Asp Cys Asn Val Ala Val Ser Gln Ile Val Asp Phe Ser Leu		
1460	1465	1470

Asp	Pro	Thr	Phe	Thr	Ile	Thr	Thr	Gln	Thr	Val	Pro	Gln	Asp	Ala
				1475					1480					1485
Val	Ser	Arg	Ser	Gln	Arg	Arg	Gly	Arg	Thr	Gly	Arg	Gly	Arg	Leu
				1490					1495					1500
Gly	Val	Tyr	Arg	Tyr	Val	Ser	Ser	Gly	Glu	Arg	Pro	Ser	Gly	Met
				1505					1510					1515
Phe	Asp	Ser	Val	Val	Leu	Cys	Glu	Cys	Tyr	Asp	Ala	Gly	Ala	Ala
				1520					1525					1530
Trp	Tyr	Glu	Leu	Thr	Pro	Ala	Glu	Thr	Thr	Val	Arg	Leu	Arg	Ala
				1535					1540					1545
Tyr	Phe	Asn	Thr	Pro	Gly	Leu	Pro	Val	Cys	Gln	Asp	His	Leu	Glu
				1550					1555					1560
Phe	Trp	Glu	Ala	Val	Phe	Thr	Gly	Leu	Thr	His	Ile	Asp	Ala	His
				1565					1570					1575
Phe	Leu	Ser	Gln	Thr	Lys	Gln	Gly	Gly	Glu	Asn	Phe	Ala	Tyr	Leu
				1580					1585					1590
Thr	Ala	Tyr	Gln	Ala	Thr	Val	Cys	Ala	Arg	Ala	Lys	Ala	Pro	Pro
				1595					1600					1605
Pro	Ser	Trp	Asp	Val	Met	Trp	Lys	Cys	Leu	Thr	Arg	Leu	Lys	Pro
				1610					1615					1620
Thr	Leu	Thr	Gly	Pro	Thr	Pro	Leu	Leu	Tyr	Arg	Leu	Gly	Ala	Val
				1625					1630					1635
Thr	Asn	Glu	Val	Thr	Leu	Thr	His	Pro	Val	Thr	Lys	Tyr	Ile	Ala
				1640					1645					1650
Thr	Cys	Met	Gln	Ala	Asp	Leu	Glu	Ile	Met	Thr	Ser	Ser	Trp	Val
				1655					1660					1665
Leu	Ala	Gly	Gly	Val	Leu	Ala	Ala	Val	Ala	Ala	Tyr	Cys	Leu	Ala
				1670					1675					1680
Thr	Gly	Cys	Ile	Ser	Ile	Ile	Gly	Arg	Leu	His	Leu	Asn	Asp	Arg

1685	1690	1695
Val Val Val Ala Pro Asp Lys Glu Ile Leu Tyr Glu Ala Phe Asp		
1700	1705	1710
Glu Met Glu Glu Cys Ala Ser Lys Ala Ala Leu Ile Glu Glu Gly		
1715	1720	1725
Gln Arg Met Ala Glu Met Leu Lys Ser Lys Ile Gln Gly Leu Leu		
1730	1735	1740
Gln Gln Ala Thr Arg Gln Ala Gln Asp Ile Gln Pro Ala Ile Gln		
1745	1750	1755
Ser Ser Trp Pro Lys Leu Glu Gln Phe Trp Ala Lys His Met Trp		
1760	1765	1770
Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu		
1775	1780	1785
Pro Gly Asn Pro Ala Val Ala Ser Met Met Ala Phe Ser Ala Ala		
1790	1795	1800
Leu Thr Ser Pro Leu Pro Thr Ser Thr Thr Ile Leu Leu Asn Ile		
1805	1810	1815
Met Gly Gly Trp Leu Ala Ser Gln Ile Ala Pro Pro Ala Gly Ala		
1820	1825	1830
Thr Gly Phe Val Val Ser Gly Leu Val Gly Ala Ala Val Gly Ser		
1835	1840	1845
Ile Gly Leu Gly Lys Ile Leu Val Asp Val Leu Ala Gly Tyr Gly		
1850	1855	1860
Ala Gly Ile Ser Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly		
1865	1870	1875
Glu Lys Pro Thr Val Glu Asp Val Val Asn Leu Leu Pro Ala Ile		
1880	1885	1890
Leu Ser Pro Gly Ala Leu Val Val Gly Val Ile Cys Ala Ala Ile		
1895	1900	1905

Leu	Arg	Arg	His	Val	Gly	Gln	Gly	Glu	Gly	Ala	Val	Gln	Trp	Met
				1910						1915				1920
Asn	Arg	Leu	Ile	Ala	Phe	Ala	Ser	Arg	Gly	Asn	His	Val	Ala	Pro
				1925						1930				1935
Thr	His	Tyr	Val	Val	Glu	Ser	Asp	Ala	Ser	Gln	Arg	Val	Thr	Gln
				1940						1945				1950
Val	Leu	Ser	Ser	Leu	Thr	Ile	Thr	Ser	Leu	Leu	Arg	Arg	Leu	His
				1955						1960				1965
Ala	Trp	Ile	Thr	Glu	Asp	Cys	Pro	Val	Pro	Cys	Ser	Gly	Ser	Trp
				1970						1975				1980
Leu	Gln	Asp	Ile	Trp	Asp	Trp	Val	Cys	Ser	Ile	Leu	Thr	Asp	Phe
				1985						1990				1995
Lys	Asn	Trp	Leu	Ser	Ser	Lys	Leu	Leu	Pro	Lys	Met	Pro	Gly	Ile
				2000						2005				2010
Pro	Phe	Ile	Ser	Cys	Gln	Lys	Gly	Tyr	Lys	Gly	Val	Trp	Ala	Gly
				2015						2020				2025
Thr	Gly	Val	Met	Thr	Thr	Arg	Cys	Pro	Cys	Gly	Ala	Asn	Ile	Ser
				2030						2035				2040
Gly	His	Val	Arg	Met	Gly	Thr	Met	Lys	Ile	Thr	Gly	Pro	Lys	Thr
				2045						2050				2055
Cys	Leu	Asn	Leu	Trp	Gln	Gly	Thr	Phe	Pro	Ile	Asn	Cys	Tyr	Thr
				2060						2065				2070
Glu	Gly	Pro	Cys	Val	Pro	Lys	Pro	Pro	Pro	Asn	Tyr	Lys	Thr	Ala
				2075						2080				2085
Ile	Trp	Arg	Val	Ala	Ala	Ser	Glu	Tyr	Val	Glu	Val	Thr	Gln	His
				2090						2095				2100
Gly	Ser	Phe	Ser	Tyr	Val	Thr	Gly	Leu	Thr	Ser	Asp	Asn	Leu	Lys
				2105						2110				2115
Val	Pro	Cys	Gln	Val	Pro	Ala	Pro	Glu	Phe	Phe	Ser	Trp	Val	Asp

	2120		2125		2130
Gly Val Gln Ile His Arg Phe Ala Pro Val Pro Gly Pro Phe Phe					
	2135		2140		2145
Arg Asp Glu Val Thr Phe Thr Val Gly Leu Asn Ser Phe Val Val					
	2150		2155		2160
Gly Ser Gln Leu Pro Cys Asp Pro Glu Pro Asp Thr Glu Val Leu					
	2165		2170		2175
Ala Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Ala Ala					
	2180		2185		2190
Ala Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Gln Ala Ser Ser					
	2195		2200		2205
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr					
	2210		2215		2220
Thr His Lys Thr Ala Tyr Asp Cys Asp Met Val Asp Ala Asn Leu					
	2225		2230		2235
Phe Met Gly Gly Asp Val Thr Arg Ile Glu Ser Asp Ser Lys Val					
	2240		2245		2250
Ile Val Leu Asp Ser Leu Asp Ser Met Thr Glu Val Glu Asp Asp					
	2255		2260		2265
Arg Glu Pro Ser Val Pro Ser Glu Tyr Leu Ile Lys Arg Arg Lys					
	2270		2275		2280
Phe Pro Pro Ala Leu Pro Pro Trp Ala Arg Pro Asp Tyr Asn Pro					
	2285		2290		2295
Val Leu Ile Glu Thr Trp Lys Arg Pro Gly Tyr Glu Pro Pro Thr					
	2300		2305		2310
Val Leu Gly Cys Ala Leu Pro Pro Thr Pro Gln Thr Pro Val Pro					
	2315		2320		2325
Pro Pro Arg Arg Arg Arg Ala Lys Val Leu Thr Gln Asp Asn Val					
	2330		2335		2340

Glu Gly Val Leu Arg Glu Met Ala Asp Lys Val Leu Ser Pro Leu	2345	2350	2355
Gln Asp Asn Asn Asp Ser Gly His Ser Thr Gly Ala Asp Thr Gly	2360	2365	2370
Gly Asp Ile Val Gln Gln Pro Ser Asp Glu Thr Ala Ala Ser Glu	2375	2380	2385
Ala Gly Ser Leu Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly	2390	2395	2400
Asp Pro Asp Leu Glu Phe Glu Pro Val Gly Ser Ala Pro Pro Ser	2405	2410	2415
Glu Gly Glu Cys Glu Val Ile Asp Ser Asp Ser Lys Ser Trp Ser	2420	2425	2430
Thr Val Ser Asp Gln Glu Asp Ser Val Ile Cys Cys Ser Met Ser	2435	2440	2445
Tyr Ser Trp Thr Gly Ala Leu Ile Thr Pro Cys Gly Pro Glu Glu	2450	2455	2460
Glu Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Met Arg Phe	2465	2470	2475
His Asn Lys Val Tyr Ser Thr Thr Ser Arg Ser Ala Ser Leu Arg	2480	2485	2490
Ala Lys Lys Val Thr Phe Asp Arg Val Gln Val Leu Asp Ala His	2495	2500	2505
Tyr Asp Ser Val Leu Gln Asp Val Lys Arg Ala Ala Ser Lys Val	2510	2515	2520
Ser Ala Arg Leu Leu Thr Val Glu Glu Ala Cys Ala Leu Thr Pro	2525	2530	2535
Pro His Ser Ala Lys Ser Arg Tyr Gly Phe Gly Ala Lys Glu Val	2540	2545	2550
Arg Ser Leu Ser Arg Arg Ala Val Asn His Ile Arg Ser Val Trp			

2555	2560	2565
Glu Asn Leu Leu Glu Asp Gln His Thr Pro Ile Asp Thr Thr Ile		
2570	2575	2580
Met Ala Lys Asn Glu Val Phe Cys Ile Asp Pro Thr Lys Gly Gly		
2585	2590	2595
Lys Lys Pro Ala Arg Leu Ile Val Tyr Pro Asp Leu Gly Val Arg		
2600	2605	2610
Val Cys Glu Lys Met Ala Leu Tyr Asp Ile Ala Gln Lys Leu Pro		
2615	2620	2625
Lys Ala Ile Met Gly Pro Ser Tyr Gly Phe Gln Tyr Ser Pro Ala		
2630	2635	2640
Glu Arg Val Asp Phe Leu Leu Lys Ala Trp Gly Ser Lys Lys Asp		
2645	2650	2655
Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val		
2660	2665	2670
Thr Glu Arg Asp Ile Arg Thr Glu Glu Ser Ile Tyr Gln Ala Cys		
2675	2680	2685
Ser Leu Pro Gln Glu Ala Arg Thr Val Ile His Ser Leu Thr Glu		
2690	2695	2700
Arg Leu Tyr Val Gly Gly Pro Met Thr Asn Ser Lys Gly Gln Ser		
2705	2710	2715
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Phe Thr Thr Ser		
2720	2725	2730
Met Gly Asn Thr Met Thr Cys Tyr Ile Lys Ala Leu Ala Ala Cys		
2735	2740	2745
Lys Ala Ala Gly Ile Val Asp Pro Val Met Leu Val Cys Gly Asp		
2750	2755	2760
Asp Leu Val Val Ile Ser Glu Ser Gln Gly Asn Glu Glu Asp Glu		
2765	2770	2775

Arg Asn Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala	2780	2785	2790
Pro Pro Gly Asp Leu Pro Arg Pro Glu Tyr Asp Leu Glu Leu Ile	2795	2800	2805
Thr Ser Cys Ser Ser Asn Val Ser Val Ala Leu Asp Ser Arg Gly	2810	2815	2820
Arg Arg Arg Tyr Phe Leu Thr Arg Asp Pro Thr Thr Pro Ile Thr	2825	2830	2835
Arg Ala Ala Trp Glu Thr Val Arg His Ser Pro Val Asn Ser Trp	2840	2845	2850
Leu Gly Asn Ile Ile Gln Tyr Ala Pro Thr Ile Trp Val Arg Met	2855	2860	2865
Val Ile Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Asp Thr	2870	2875	2880
Leu Asn Gln Asn Leu Asn Phe Glu Met Tyr Gly Ala Val Tyr Ser	2885	2890	2895
Val Asn Pro Leu Asp Leu Pro Ala Ile Ile Glu Arg Leu His Gly	2900	2905	2910
Leu Glu Ala Phe Ser Leu His Thr Tyr Ser Pro His Glu Leu Ser	2915	2920	2925
Arg Val Ala Ala Thr Leu Arg Lys Leu Gly Ala Pro Pro Leu Arg	2930	2935	2940
Ala Trp Lys Ser Arg Ala Arg Ala Val Arg Ala Ser Leu Ile Ala	2945	2950	2955
Gln Gly Ala Arg Ala Ala Ile Cys Gly Arg Tyr Leu Phe Asn Trp	2960	2965	2970
Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Leu Pro Glu Ala Ser	2975	2980	2985
Arg Leu Asp Leu Ser Gly Trp Phe Thr Val Gly Ala Gly Gly Gly			

	2990		2995		3000
Asp	Ile	Tyr	His	Ser	Val
Ser	His	Ala	Arg	Pro	Arg
Leu	Leu	Leu			
	3005		3010		3015
Leu	Cys	Leu	Leu	Leu	Ser
Val	Gly	Val	Gly	Ile	Phe
Leu	Leu				
	3020		3025		3030
Pro	Ala	Arg			
	3033				

Sequence ID No. 9
 Sequence Length: 3,033
 Sequence Type: amino acid
 Topology: linear
 Molecule Type: protein

Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn	Thr	5	10	15
Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly	Gln	Ile	20	25	30
Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro	Arg	Leu	Gly	35	40	45
Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	Pro	Arg	Gly	50	55	60
Arg	Arg	Gln	Pro	Ile	Pro	Lys	Asp	Arg	Arg	Ser	Thr	Gly	Lys	Ser	65	70	75
Trp	Gly	Lys	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	Gly	80	85	90
Cys	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	95	100	105
Thr	Trp	Gly	Pro	Thr	Asp	Pro	Arg	His	Arg	Ser	Arg	Asn	Leu	Gly	110	115	120
Arg	Val	Ile	Asp	Thr	Ile	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly	125	130	135
Tyr	Ile	Pro	Val	Val	Gly	Ala	Pro	Val	Gly	Gly	Val	Ala	Arg	Ala	140	145	150
Leu	Ala	His	Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Ile	Asn	Tyr	Ala	155	160	165

Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala		
170	175	180
Leu Leu Ser Cys Val Thr Met Pro Val Ser Ala Val Glu Val Arg		
185	190	195
Asn Ile Ser Ser Ser Tyr Tyr Ala Thr Asn Asp Cys Ser Asn Asn		
200	205	210
Ser Ile Thr Trp Gln Leu Thr Asp Ala Val Leu His Leu Pro Gly		
215	220	225
Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu Arg Cys Trp Ile		
230	235	240
Gln Val Thr Pro Asp Val Ala Val Lys His Arg Gly Ala Leu Thr		
245	250	255
Arg Ser Leu Arg Thr His Val Asp Met Ile Val Met Ala Ala Thr		
260	265	270
Ala Cys Ser Ala Leu Tyr Val Gly Asp Val Cys Gly Ala Val Met		
275	280	285
Ile Leu Ser Gln Ala Phe Met Val Ser Pro Gln Arg His Asn Phe		
290	295	300
Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly His Ile Thr Gly		
305	310	315
His Arg Met Ala Trp Asp Met Met Leu Asn Trp Ser Pro Thr Leu		
320	325	330
Ala Met Ile Leu Ala Tyr Ala Ala Arg Val Pro Glu Leu Val Leu		
335	340	345
Glu Ile Ile Phe Gly Gly His Trp Gly Val Ala Phe Gly Leu Gly		
350	355	360
Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val Val Ala Ile Leu		
365	370	375
Leu Leu Val Ala Gly Val Asp Ala Ser Thr Tyr Ser Thr Gly Gln		

380	385	390
Gln Ala Gly Arg Ala Ala Tyr Gly Ile Ser Ser Leu Phe Asn Thr		
395	400	405
Gly Ala Lys Gln Asn Leu His Leu Ile Asn Thr Asn Gly Ser Trp		
410	415	420
His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser Leu Glu Thr		
425	430	435
Gly Phe Ile Ala Ser Leu Val Tyr Tyr Arg Arg Phe Asn Ser Ser		
440	445	450
Gly Cys Pro Glu Arg Leu Ser Ser Cys Arg Gly Leu Asp Asp Phe		
455	460	465
Arg Ile Gly Trp Gly Thr Leu Glu Tyr Glu Thr Asn Val Thr Asn		
470	475	480
Asp Glu Asp Met Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg Pro		
485	490	495
Cys Gly Ile Val Pro Ala Arg Thr Val Cys Gly Pro Val Tyr Cys		
500	505	510
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Lys Gln Gly		
515	520	525
Val Pro Thr Tyr Thr Trp Gly Glu Asn Glu Thr Asp Val Phe Leu		
530	535	540
Leu Asn Ser Thr Arg Pro Pro Arg Gly Ala Trp Phe Gly Cys Thr		
545	550	555
Trp Met Asn Gly Thr Gly Phe Thr Lys Thr Cys Gly Ala Pro Pro		
560	565	570
Cys Arg Ile Arg Lys Asp Tyr Asn Ser Thr Ile Asp Leu Leu Cys		
575	580	585
Pro Thr Asp Cys Phe Arg Lys His Pro Asp Ala Thr Tyr Leu Lys		
590	595	600

Cys Gly Ala Gly Pro Trp Leu Thr Pro Arg Cys Leu Val Asp Tyr		
605	610	615
Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe Thr Ile		
620	625	630
Phe Lys Ala Arg Met Tyr Val Gly Gly Val Glu His Arg Phe Ser		
635	640	645
Ala Ala Cys Asn Phe Thr Arg Gly Asp Arg Cys Arg Leu Glu Asp		
650	655	660
Arg Asp Arg Gly Gln Gln Ser Pro Leu Leu His Ser Thr Thr Glu		
665	670	675
Trp Ala Val Phe Pro Cys Ser Phe Ser Asp Leu Pro Ala Leu Ser		
680	685	690
Thr Gly Leu Leu His Leu His Gln Asn Ile Val Asp Val Gln Tyr		
695	700	705
Leu Tyr Gly Leu Ser Pro Ala Leu Thr Arg Tyr Ile Val Lys Trp		
710	715	720
Glu Trp Val Ile Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val		
725	730	735
Cys Ala Cys Leu Trp Met Leu Asn Ile Leu Gly Gln Ala Glu Ala		
740	745	750
Ala Leu Glu Lys Leu Ile Ile Leu His Ser Ala Ser Ala Ala Ser		
755	760	765
Ala Asn Gly Pro Leu Trp Phe Phe Ile Phe Phe Thr Ala Ala Trp		
770	775	780
Tyr Leu Lys Gly Arg Val Val Pro Val Ala Thr Tyr Ser Val Leu		
785	790	795
Gly Leu Trp Ser Phe Leu Leu Leu Val Leu Ala Leu Pro Gln Gln		
800	805	810
Ala Tyr Ala Leu Asp Ala Ala Glu Gln Gly Glu Leu Gly Leu Ala		

815	820	825
Ile Leu Val Ile Ile Ser Ile Phe Thr	Leu Thr Pro Ala Tyr Lys	
830	835	840
Ile Leu Leu Ser Arg Ser Val Trp Trp	Leu Ser Tyr Met Leu Val	
845	850	855
Leu Ala Glu Ala Gln Ile Gln Gln Trp	Val Pro Pro Leu Glu Val	
860	865	870
Arg Gly Gly Arg Asp Gly Ile Ile Trp	Val Ala Val Ile Leu His	
875	880	885
Pro Arg Leu Val Phe Glu Val Thr Lys	Trp Leu Leu Ala Ile Leu	
890	895	900
Gly Pro Ala Tyr Leu Leu Arg Ala Ser	Leu Leu Arg Ile Pro Tyr	
905	910	915
Phe Val Arg Ala His Ala Leu Leu Arg	Val Cys Thr Leu Val Lys	
920	925	930
His Leu Ala Gly Ala Arg Tyr Ile Gln	Met Leu Leu Ile Thr Ile	
935	940	945
Gly Arg Trp Thr Gly Thr Tyr Ile Tyr	Asp His Leu Ser Pro Leu	
950	955	960
Ser Thr Trp Ala Ala Gln Gly Leu Arg	Asp Leu Ala Ile Ala Val	
965	970	975
Glu Pro Val Val Phe Ser Pro Met Glu	Lys Lys Val Ile Val Trp	
980	985	990
Gly Ala Glu Thr Val Ala Cys Gly Asp	Ile Leu His Gly Leu Pro	
995	1000	1005
Val Ser Ala Arg Leu Gly Arg Glu Val	Leu Leu Gly Pro Ala Asp	
1010	1015	1020
Gly Tyr Thr Ser Lys Gly Trp Asn Leu	Leu Ala Pro Ile Thr Ala	
1025	1030	1035

Tyr Thr Gln Gln Thr Arg Gly Leu Leu Gly Ala Ile Val Val Ser		
1040	1045	1050
Leu Thr Gly Arg Asp Lys Asn Glu Gln Ala Gly Gln Val Gln Val		
1055	1060	1065
Leu Ser Ser Val Thr Gln Thr Phe Leu Gly Thr Ser Ile Ser Gly		
1070	1075	1080
Val Leu Trp Thr Val Tyr His Gly Ala Gly Asn Lys Thr Leu Ala		
1085	1090	1095
Gly Pro Lys Gly Pro Val Thr Gln Met Tyr Thr Ser Ala Glu Gly		
1100	1105	1110
Asp Leu Val Gly Trp Pro Ser Pro Pro Gly Thr Lys Ser Leu Asp		
1115	1120	1125
Pro Cys Thr Cys Gly Ala Val Asp Leu Tyr Leu Val Thr Arg Asn		
1130	1135	1140
Ala Asp Val Ile Pro Val Arg Arg Lys Asp Asp Arg Arg Gly Ala		
1145	1150	1155
Leu Leu Ser Pro Arg Pro Leu Ser Thr Leu Lys Gly Ser Ser Gly		
1160	1165	1170
Gly Pro Val Leu Cys Ser Arg Gly His Ala Val Gly Leu Phe Arg		
1175	1180	1185
Ala Ala Val Cys Ala Arg Gly Val Ala Lys Ser Ile Asp Phe Ile		
1190	1195	1200
Pro Val Glu Ser Leu Asp Ile Ala Thr Arg Thr Pro Ser Phe Ser		
1205	1210	1215
Asp Asn Ser Ala Pro Pro Ala Val Pro Gln Ser Tyr Gln Val Gly		
1220	1225	1230
Tyr Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro		
1235	1240	1245
Ala Ala Tyr Ala Ser Gln Gly Tyr Lys Val Leu Val Leu Asn Pro		

1250	1255	1260
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala		
1265	1270	1275
His Gly Ile Asn Pro Asn Ile Arg Thr Gly Val Arg Thr Val Thr		
1280	1285	1290
Thr Gly Asp Ser Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Ile Ala		
1295	1300	1305
Asp Gly Gly Cys Ala Ala Gly Ala Tyr Asp Ile Ile Ile Cys Asp		
1310	1315	1320
Glu Cys His Ser Val Asp Ala Thr Thr Ile Leu Gly Ile Gly Thr		
1325	1330	1335
Val Leu Asp Gln Ala Glu Thr Ala Gly Val Arg Leu Val Val Leu		
1340	1345	1350
Ala Thr Ala Thr Pro Pro Gly Thr Val Thr Thr Pro His Ser Asn		
1355	1360	1365
Ile Glu Glu Val Ala Leu Gly His Glu Gly Glu Ile Pro Phe Tyr		
1370	1375	1380
Gly Lys Ala Ile Pro Leu Ala Phe Ile Lys Gly Gly Arg His Leu		
1385	1390	1395
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Ala		
1400	1405	1410
Leu Arg Gly Thr Gly Val Asn Ala Val Ala Tyr Tyr Arg Gly Leu		
1415	1420	1425
Asp Val Ser Val Ile Pro Thr Gln Gly Asp Val Val Val Val Ala		
1430	1435	1440
Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val		
1445	1450	1455
Ile Asp Cys Asn Val Ala Val Ser Gln Ile Val Asp Phe Ser Leu		
1460	1465	1470

Asp Pro Thr Phe Thr Ile Thr Thr Gln Thr Val Pro Gln Asp Ala		
1475	1480	1485
Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Leu		
1490	1495	1500
Gly Ile Tyr Arg Tyr Val Ser Ser Gly Glu Gly Pro Ser Gly Met		
1505	1510	1515
Phe Asp Ser Val Val Pro Cys Glu Cys Tyr Asp Ala Gly Ala Ala		
1520	1525	1530
Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala		
1535	1540	1545
Tyr Phe Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu		
1550	1555	1560
Phe Trp Glu Ala Val Phe Thr Gly Leu Thr His Ile Asn Ala His		
1565	1570	1575
Phe Leu Ser Gln Thr Lys Gln Gly Gly Glu Asn Phe Ala Tyr Leu		
1580	1585	1590
Thr Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Lys Ala Pro Pro		
1595	1600	1605
Pro Ser Trp Asp Val Met Trp Lys Cys Leu Thr Arg Leu Lys Pro		
1610	1615	1620
Thr Leu Thr Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val		
1625	1630	1635
Thr Asn Glu Val Thr Leu Thr His Pro Val Thr Lys Tyr Ile Ala		
1640	1645	1650
Thr Cys Met Gln Ala Asp Leu Glu Ile Met Thr Ser Ser Trp Val		
1655	1660	1665
Leu Ala Gly Gly Val Leu Ala Ala Val Ala Ala Tyr Cys Leu Ala		
1670	1675	1680
Thr Gly Cys Ile Ser Ile Ile Gly Arg Leu His Leu Asn Asp Arg		

1685	1690	1695
Val Val Val Thr Pro Asp Lys Glu Ile Leu Tyr Glu Ala Phe Asp		
1700	1705	1710
Glu Met Glu Glu Cys Ala Ser Lys Ala Ala Leu Ile Glu Glu Gly		
1715	1720	1725
Gln Arg Met Ala Glu Met Leu Lys Ser Lys Ile Gln Gly Leu Leu		
1730	1735	1740
Gln Gln Ala Thr Arg Gln Ala Gln Gly Met Gln Pro Ala Ile Gln		
1745	1750	1755
Ser Ser Trp Pro Lys Leu Glu Gln Phe Trp Ala Lys His Met Trp		
1760	1765	1770
Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu		
1775	1780	1785
Pro Gly Asn Pro Ala Val Ala Ser Met Met Ala Phe Ser Ala Ala		
1790	1795	1800
Leu Thr Ser Pro Leu Pro Thr Ser Thr Thr Ile Leu Leu Asn Ile		
1805	1810	1815
Met Gly Gly Trp Leu Ala Ser Gln Ile Ala Pro Pro Ala Gly Ala		
1820	1825	1830
Thr Gly Phe Val Val Ser Gly Leu Val Gly Ala Ala Val Gly Ser		
1835	1840	1845
Ile Gly Leu Gly Lys Ile Leu Val Asp Val Leu Ala Gly Tyr Gly		
1850	1855	1860
Ala Gly Ile Ser Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly		
1865	1870	1875
Glu Lys Pro Thr Val Glu Asp Val Val Asn Leu Leu Pro Ala Ile		
1880	1885	1890
Leu Ser Pro Gly Ala Leu Val Val Gly Val Ile Cys Ala Ala Ile		
1895	1900	1905

Leu Arg Arg His Val Gly Gln Gly Glu Gly Ala Val Gln Trp Met		
1910	1915	1920
Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ala Pro		
1925	1930	1935
Thr His Tyr Val Val Glu Ser Asp Ala Ser Gln Arg Val Thr Gln		
1940	1945	1950
Val Leu Ser Ser Leu Thr Ile Thr Ser Leu Leu Arg Arg Leu His		
1955	1960	1965
Ala Trp Ile Thr Glu Asp Cys Pro Ile Pro Cys Ser Gly Ser Trp		
1970	1975	1980
Leu Gln Asp Ile Trp Asp Trp Val Cys Ser Ile Leu Thr Asp Phe		
1985	1990	1995
Lys Asn Trp Leu Ser Ser Lys Leu Leu Pro Lys Met Pro Gly Ile		
2000	2005	2010
Pro Phe Ile Ser Cys Gln Lys Gly Tyr Lys Gly Val Trp Ala Gly		
2015	2020	2025
Thr Gly Val Met Thr Thr Arg Tyr Pro Cys Gly Ala Asn Ile Ser		
2030	2035	2040
Gly His Val Arg Met Gly Thr Met Lys Ile Thr Gly Pro Lys Thr		
2045	2050	2055
Cys Leu Asn Leu Trp Gln Gly Thr Phe Pro Ile Asn Cys Tyr Thr		
2060	2065	2070
Glu Gly Pro Cys Val Pro Lys Pro Pro Pro Asn Tyr Lys Thr Ala		
2075	2080	2085
Ile Trp Arg Val Ala Ala Ser Glu Tyr Val Glu Val Thr Gln His		
2090	2095	2100
Gly Ser Phe Ser Tyr Val Thr Gly Leu Thr Ser Asp Asn Leu Lys		
2105	2110	2115
Val Pro Cys Gln Val Pro Ala Pro Glu Phe Phe Ser Trp Val Asp		

2120	2125	2130
Gly Val Gln Ile His Arg Phe Ala Pro Val Pro Gly Pro Phe Phe		
2135	2140	2145
Arg Asp Glu Val Thr Phe Thr Val Gly Leu Asn Ser Phe Val Val		
2150	2155	2160
Gly Ser Gln Leu Pro Cys Asp Pro Glu Pro Asp Thr Glu Val Leu		
2165	2170	2175
Ala Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu Ala Ala		
2180	2185	2190
Ala Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Gln Ala Ser Ser		
2195	2200	2205
Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr		
2210	2215	2220
Thr His Lys Thr Ala Tyr Asp Cys Asp Met Val Asp Ala Asn Leu		
2225	2230	2235
Phe Met Gly Gly Asp Val Thr Arg Ile Glu Ser Asp Ser Lys Val		
2240	2245	2250
Ile Val Leu Asp Ser Leu Asp Ser Met Thr Glu Val Glu Asp Asp		
2255	2260	2265
Arg Glu Pro Ser Val Pro Ser Glu Tyr Leu Ile Lys Arg Arg Lys		
2270	2275	2280
Phe Pro Pro Ala Leu Pro Pro Trp Ala Arg Pro Asp Tyr Asn Pro		
2285	2290	2295
Val Leu Ile Glu Thr Trp Lys Arg Pro Gly Tyr Glu Pro Pro Thr		
2300	2305	2310
Val Leu Gly Cys Ala Leu Pro Pro Thr Leu Gln Thr Pro Val Pro		
2315	2320	2325
Pro Pro Arg Arg Arg Arg Ala Lys Ile Leu Thr Gln Asp Asp Val		
2330	2335	2340

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number:

0 532 167 A3

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: **92306952.0**

(51) Int. Cl.⁵ **C12N 15/51, A61K 39/29,
G01N 33/576, C07K 15/00,
C12P 21/08**

(22) Date of filing: **30.07.92**

(30) Priority: **09.08.91 JP 287402/91
05.12.91 JP 360441/91**

(43) Date of publication of application:
17.03.93 Bulletin 93/11

(84) Designated Contracting States:
BE CH DE DK FR GB IT LI NL

(36) Date of deferred publication of the search report:
31.03.93 Bulletin 93/13

(71) Applicant: **IMMUNO JAPAN INC.**
**4-28-14-701, Ogikubo Suginami-Ku
Tokyo 167(JP)**

(72) Inventor: **Okamoto, Hiroaki**
3132-24, Yakushiji, Minami Kawachi Machi
Kawachi-gun; Tochigi-ken 329-04(JP)
Inventor: **Nakamura, Tetsuo**
14-701, 4-chome 28, Ogikubo
Suginami-ku, Tokyo 167(JP)

(74) Representative: **Arthur, Bryan Edward**
4 Dyers Buildings Holborn
London, EC1N 2JT (GB)

(54) **Non-A, Non-B Hepatitis virus genome, polynucleotides, polypeptides, antigen, antibody and detection systems.**

(57) **Non-A, non-B hepatitis (NANB hepatitis) virus RNA and its corresponding polypeptide, related antigen, antibody, and detection systems for detecting NANB hepatitis antigen or antibodies.**

EP 0 532 167 A3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number

EP 92 30 6952

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
X,P	WO-A-9 114 779 (MITSUI TOATSU CHEMICALS, INC.) * figures 1-9 * ---	1-13	C12N15/51 A61K39/29 G01N33/576 C07K15/00 C12P21/08
X,P	EP-A-0 468 657 (TONEN CORPORATION) * the whole document * ---	1-13	
X,P	EP-A-0 485 209 (IMMUNO JAPAN INC) * the whole document * ---	1-13	
E	WO-A-9 219 743 (CHIRON CORPORATION) * the whole document * ---	1-13	
E	EP-A-0 516 859 (TORAY INDUSTRIES INC) * the whole document * ---	6-13	
X	BIOCHEM. BIOPHYS. RES. COMMUN. vol. 170, no. 3, 1990, pages 1021 - 1025 N. ENOMOTO ET AL. 'there are two major types of hepatitis C virus in Japan' * the whole document * -----	1-13	
			TECHNICAL FIELDS SEARCHED (Int. Cl.5)
			C07K A61K C12N C12Q G01N
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 01 FEBRUARY 1993	Examiner SKELLY J.M.
CATEGORY OF CITED DOCUMENTS		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons ----- & : member of the same patent family, corresponding document	
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document			

1
EPO FORM 1303 (12.92) (P0601)

Glu Gly Ile Leu Arg Glu Met Ala Asp Lys Val Leu Ser Pro Leu	2345	2350	2355
Gln Asp Asn Asn Asp Ser Gly His Ser Thr Gly Ala Asp Thr Gly	2360	2365	2370
Gly Asp Ile Val Gln Gln Pro Ser Asp Glu Thr Ala Ala Ser Glu	2375	2380	2385
Ala Gly Ser Leu Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly	2390	2395	2400
Asp Pro Asp Leu Glu Phe Glu Pro Val Gly Ser Ala Pro Pro Ser	2405	2410	2415
Glu Gly Glu Cys Glu Val Ile Asp Ser Asp Ser Lys Ser Trp Ser	2420	2425	2430
Thr Val Ser Asp Gln Glu Asp Ser Val Ile Cys Cys Ser Met Ser	2435	2440	2445
Tyr Ser Trp Thr Gly Ala Leu Ile Thr Pro Cys Gly Pro Glu Glu	2450	2455	2460
Glu Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Met Arg Phe	2465	2470	2475
His Asn Lys Val Tyr Ser Thr Thr Ser Arg Ser Ala Ser Leu Arg	2480	2485	2490
Ala Lys Lys Val Thr Phe Asp Arg Val Gln Val Leu Asp Ala His	2495	2500	2505
Tyr Asp Ser Val Leu Gln Asp Val Lys Arg Ala Ala Ser Lys Val	2510	2515	2520
Gly Ala Arg Leu Leu Thr Val Glu Glu Ala Cys Ala Leu Thr Pro	2525	2530	2535
Pro His Ser Ala Lys Ser Arg Tyr Gly Phe Gly Ala Lys Glu Val	2540	2545	2550
Arg Ser Leu Ser Arg Arg Ala Val Asn His Ile Arg Ser Val Trp			

2555	2560	2565
Glu Asn Leu Leu Glu Asp Gln Arg Thr Pro Ile Asp Thr Thr Ile		
2570	2575	2580
Met Ala Lys Asn Glu Val Phe Cys Ile Asp Pro Thr Lys Gly Gly		
2585	2590	2595
Lys Lys Pro Ala Arg Leu Ile Val Tyr Pro Asp Leu Gly Val Arg		
2600	2605	2610
Val Cys Glu Lys Met Ala Leu Tyr Asp Ile Thr Gln Lys Leu Pro		
2615	2620	2625
Lys Ala Ile Met Gly Pro Ser Tyr Gly Phe Gln Tyr Ser Pro Ala		
2630	2635	2640
Glu Arg Val Asp Phe Leu Leu Lys Ala Trp Gly Ser Lys Lys Asp		
2645	2650	2655
Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val		
2660	2665	2670
Thr Glu Arg Asp Ile Arg Thr Glu Glu Ser Ile Tyr Gln Ala Cys		
2675	2680	2685
Ser Leu Pro Gln Glu Ala Arg Thr Val Ile His Ser Leu Thr Glu		
2690	2695	2700
Arg Leu Tyr Val Gly Gly Pro Met Thr Asn Ser Lys Gly Gln Ser		
2705	2710	2715
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Phe Thr Thr Ser		
2720	2725	2730
Met Gly Asn Thr Met Thr Cys Tyr Ile Lys Ala Leu Ala Ala Cys		
2735	2740	2745
Lys Ala Ala Gly Ile Val Asp Pro Val Met Leu Val Cys Gly Asp		
2750	2755	2760
Asp Leu Val Val Ile Ser Glu Ser Gln Gly Asn Glu Glu Asp Glu		
2765	2770	2775

Arg Asn Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala	2780	2785	2790
Pro Pro Gly Asp Leu Pro Arg Pro Glu Tyr Asp Leu Glu Leu Ile	2795	2800	2805
Thr Ser Cys Ser Ser Asn Val Ser Val Ala Leu Asp Ser Arg Gly	2810	2815	2820
Arg Arg Arg Tyr Phe Leu Thr Arg Asp Pro Thr Thr Pro Ile Thr	2825	2830	2835
Arg Ala Ala Trp Glu Thr Val Arg His Ser Pro Val Asn Ser Trp	2840	2845	2850
Leu Gly Asn Ile Ile Gln Tyr Ala Pro Thr Ile Trp Val Arg Met	2855	2860	2865
Val Ile Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Asp Thr	2870	2875	2880
Leu Asn Gln Asn Leu Asn Phe Glu Met Tyr Gly Ala Val Tyr Ser	2885	2890	2895
Val Asn Pro Leu Asp Leu Pro Ala Ile Ile Glu Arg Leu His Gly	2900	2905	2910
Leu Glu Ala Phe Ser Leu His Thr Tyr Ser Pro His Glu Leu Ser	2915	2920	2925
Arg Val Ala Ala Thr Leu Arg Lys Leu Gly Ala Pro Pro Leu Arg	2930	2935	2940
Ala Trp Lys Ser Arg Ala Arg Ala Val Arg Ala Ser Leu Ile Ala	2945	2950	2955
Gln Gly Ala Arg Ala Ala Ile Cys Gly Arg Tyr Leu Phe Asn Trp	2960	2965	2970
Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Leu Pro Glu Ala Ser	2975	2980	2985
Arg Leu Asp Leu Ser Gly Trp Phe Thr Val Gly Ala Gly Gly Gly			

